From Dataveillance to Data Economy: Firm View on Data Protection

Thesis

Business School

From Dataveillance to Data Economy

Firm View on Data Protection

Sara Degli Esposti

BA Sociology (Hons), MSc Business Administration and Quantitative Methods

Thesis submitted in accordance with the requirements of

The Open University for the degree of Doctor of Philosophy

1st of June 2016

From Dataveillance to Data Economy

Firm View on Data Protection

Sara Degli Esposti

BA in Sociology (Hons), MSc in Business Administration and Quantitative Methods

Thesis submitted in accordance with the requirements of

The Open University for the degree of Doctor of Philosophy

Submission date: 30th of September 2015

Viva voce examination date: 16th of February 2016

Submission date of the final version with minor revision: 9th of May 2016

Department of People and Organisations – The Open University

Walton Hall, Milton Keynes, MK7 6AA, United Kingdom

Suggested citation:

Degli-Esposti, Sara (2016), *From Dataveillance to Data Economy: Firm View on Data Protection*, PhD thesis, 1st of June 2016, The Open University Business School, Milton Keynes (UK), pp. 329.

Abstract

The increasing availability of electronic records and the expanded reliance on online communications and services have made available a huge amount of data about people's behaviours, characteristics, and preferences. Advancements in data processing technology, known as *big data*, offer opportunities to increase organisational efficiency and competitiveness. Analytically sophisticated companies excel in their ability to extract value from the analysis of digital data. However, in order to exploit the potential economic benefits produced by big data and analytics, issues of data privacy and information security need to be addressed. In Europe, organisations processing personal data are being required to implement basic data protection principles, which are considered difficult to implement in big data environments. Little is known in the privacy studies literature about how companies manage the trade-off between data usage and data protection. This study contributes to explore the corporate data privacy environment, by focusing on the interrelationship between the data protection legal regime, the application of big data analytics to achieve corporate objectives, and the creation of an organisational privacy culture. It also draws insights from surveillance studies, particularly the idea of dataveillance, to identify potential limitations of the current legal privacy regime. The findings from the analysis of survey data show that big data and data protection support each other, but also that some frictions can emerge around data collection and data fusion. The demand for the integration of different data sources poses challenges to the implementation of data protection principles. However, this study finds no evidence that data protection laws prevent data gathering. Implications relevant for the debate on the reform of European data protection law are also drawn from these findings.

Key words: Data Protection Law; Information Privacy; Dataveillance; Information Security.

Acknowledgements

I would like to thank all the people who have contributed to this incredible journey.

My deepest gratitude goes, in particular, to my PhD supervisors for their patience, active support and encouragement during all stages of this project. Kirstie Ball, Liz Daniel, Maureen Meadows and, during the first year, Luciano Batista, helped me believe in myself and grow as an independent researcher. Our supervision meetings have been great examples of professionalism, integrity, and team cooperation.

My appreciation also goes to the OU administrative staff, especially to Jackie Fry, Chris Davey, Debby Hing, and Su Prior from the Research School. I have also benefitted greatly from suggestions given by my third-party monitor, Nik Winchester. A special thanks to all my colleagues from the Surprise project, especially Sally Dibb.

Many people helped me develop this PhD project. Special thanks go to Daniel Nagel for his comments and support; Alessandro Acquisti and Chris Hoofnagle for their suggestions; David Lyon, David Murakami-Wood, Colin Bennett, Kelly Gates, and all the Surveillance Studies Network community for their critical insights. I would also like to express my gratitude to Caspar Bowden, anti-surveillance campaigner, who passed away on July 9, 2015, for sharing with me his reflections on communication anonymity and the complexity of digital surveillance.

I would like to thank Patricia Junkes, Gianluca D'Antonio, Stefano Ciminelli, and Andrea Moro, for revising early versions of the questionnaire; my appreciation also goes to those members of *The National Association of Data Protection Officers* (NAPDO), like Martin Hoskins and Rowenna Fielding, who gave their feedback on the final version of the survey.

I would like to thank those who contributed to give visibility to the study as well as to invite professionals to participate in the study: Carl Wiper from the *Information Commissioner's Office* (ICO); Chris Tiernan and Jon Hall of *British Computer Society Effective Leadership in IT* (ELITE) *Group*; Omer Tene and Jedidiah Bracy of the *International Association of Privacy Professiona*ls

(IAPP); Massimo Attoresi of the *European Data Protection Supervisor* (EDPS) authority; Ricard Martínez Martínez of *Asociación Profesional Española de Privacidad* (APEP); Llorenç Pagés Casas of *Asociación de técnicos de informática* (ATI); Stewart Dresner and Laura Linkomies of *Privacy Laws & Business*; Gavin Blackett of the Operation Research (OR) Society.

Special thanks go to Chris Tiernan and Jon Hall, who provided constructive support and guidance in the revision process of the BigDP Study Report.

I am indebted to my colleagues – thanks to Letsema Mbayi, Ziad El Otell, Camilla Cerutti and Mauro, Tamim Elbasha, Paul Grayson – and to my dear friends Lucia Debertol and Pinelopi Troullinou.

I gratefully acknowledge the 'The New Transparency Project: Surveillance and Social Sorting', funded by the *Social Sciences and Humanities Research Council* of Canada, which as sponsored this study with a doctoral studentship.

This dissertation is dedicated to my husband, who has always believed in my potential as a researcher, and to my son, who helped me increase dramatically the productivity of my working hours. In memory of my mother-in-law, whose spiritual strength and determination have profoundly influenced the decisions I took in 2010, which have made possible being here today.

Table of Contents

List of Figures

List of Tables

CHAPTER ONE

Introduction

1.1 Introduction

This chapter presents the background of the research and states the research problem while also justifying the decision of adopting a specific research design. Explanations of the importance of investigating the topic of data protection and big data analytics from a business studies perspective are, then, provided, followed by a brief discussion of the intended contribution to academic and business knowledge. The chapter ends with a clear delimitation of the scope of the study and a presentation of the structure of the thesis.

1.2 Research rationale: Data protection law and big data technology

Data produced by internet users are growing so fast that it has been estimated that we are already unable to store all the digital information we produce (Gantz and Reinsel 2010). Any Internet-mediated transaction creates some kind of digital footprint. The creation of digital contents, metadata and electronic records is a global phenomenon: seventy *percent* of all digital data produced in 2009 was stored in the Western world, namely between North America and Europe, though in 2010 China became the largest producer of personal location data, thanks to its 800 million mobile phones in use (TheEconomist 2010).

Data are managed and accessed by means of sophisticated information management systems whose evolution has been indicated with the term 'big data' in recent times. The term 'big data' is meant to describe a universe of very large datasets that hold a variety of data types (Cavoukian, Stewart et al. 2014). The novelty of big data can be identified in the data retrieval and processing challenges of brought by the exponential accumulation of data and the presence of unstructured data collected in real time (Madden 2012). Big data platforms mostly rely on distributed storage

technology rather than local storage (Hashem, Yaqoob et al. 2015). By accessing data located in different silos or servers, big data technology allows analysts to run complex queries and to statistically analyse data to identify common patterns and make predictions. The implementation of this technology is of monetary value. Big data and their analysis is considered by many the oil of the digital economy (Davenport, Barth et al. 2012).

Organisational ability to leverage data to achieve business objectives, such as market penetration or expansion, represents a fundamental source of competitive advantage (Bell 2015). Companies which outperform competitors through the constant analysis of very large datasets are called *analytical competitors* (Davenport and Harris 2007, Davenport, Harris et al. 2010, Davenport 2014). This type of organisation invests heavily in information technologies and employs data analysts capable of identifying solutions to improve organisational efficiency, competitiveness and innovative potential (Davenport, Harris et al. 2010). Some firms even generate revenues from the collection, assemblage, sale, and analysis of data. The new data economy not only contributes to the overall digital economy (Naone 2008), but it also allows organisations to create new personalised services, recommendation systems and to anticipate emerging trends (Manyika, Chui et al. 2011).

Individual information is extremely valuable to organisations that monitor people's activities in search of business insights. Consumers are particularly exposed to data monitoring as their actions are increasingly visible and easy to track (Lace 2005, Turow and Draper 2012). Since data, however, refer quite often to identifiable persons, the safeguard of people's privacy and data integrity becomes strictly intertwined with considerations related to data usage and exploitation (Milne 2000). A number of laws and regulations have been enacted across developed countries to force public and private organisations to comply with various information management principles in the attempt to protect data privacy. Since data represent a key asset, they need to be protected for the sake of both individuals and organisations.

In Europe, privacy has not only been recognised as a fundamental human value, safeguarded under comprehensive legislation (CoE 1981, EC/46 1995, Bignami 2007), but also *everyone has the right to the protection of personal data concerning him or her* (EU 2000, EU 2007). To ensure the fulfilment of this principle, several initiatives have been undertaken. Organisations which handle personal data have to comply with data protection principles stated in the EU Data Protection Directive 1995 (EC/46 1995). According to these principles, data-subjects, defined as the person the data refer to, have the right: (1) to be informed about the collection and use of their personal data; (2) to deny or grant consent to the collection and processing of their personal data; (3) to access their personal data; and (4) to object to the processing of their personal data. Data-controllers or processors, which are organisations collecting and processing individuals' data, have the duty: (1) of ensuring that data are accurate; (2) of clarifying the purpose for which data are collected and (3) of using data only for that purpose; (4) of retaining data only until the objective for which they have been collected has been achieved; (5) of securing data; (6) of being accountable for the respect of the principles already stated.

While data protection regulation sets the standards of lawful use of data, independent data protection agencies are in charge of ensuring compliance. As transborder data flow to countries with an inadequate level of protection has been forbidden, multilateral agreements (US-EU 2000) and legal instruments, such as Binding Corporate Rules (Kong 2010) have been enacted to guarantee safe data migration. The peculiarity of the electronic communication sector, with respect to this matter, has also been acknowledged and specific measures have been taken (EC/58 2002, EC/136 2009). Legislative attempts to protect data by means of stricter privacy regulations have often been criticised for adding excessive burdens on companies (Samiee 1984). On the contrary, any initiative of harmonization of data protection practices and principles was intended to facilitate the free flow of personal data across frontiers (EC 1995), and data protection law was never meant to hamper the development of the ICT sector or to cause disruption in the banking and insurance sectors (OECD 1980, EC 1995). Nevertheless, the problem of how to increase organisational information privacy and security procedures remains because the more

often data about individuals are collected and stored, the more likely data theft and loss becomes (Mulligan and Perzanowski 2007). Furthermore, enacting data protection principles is especially challenging in the context of massive data analysis also known as big data analytics (Cate, Cullen et al. 2013).

Data protection principles have been criticised for being outdated and unable to safeguard information privacy in current times characterised by user-generated content, massive data collection and analysis, and minimal storage costs (Bamberger and Mulligan 2011). The 'consent' and 'purpose limitation' provisions have been especially criticised for clashing with the reality of both users and organisations. Users of online services give their consent to privacy policies impossible to read (Milne, Culnan et al. 2006), while the current regulatory regime does not stop organisations from deploying increasingly sophisticated tools to track consumers online and offline (Tene and Polenetsky 2012).

These considerations have produced an ongoing debate between the critics and the supporters of data protection principles and several attempts to change the current regulatory landscape (Tene 2010). Despite huge efforts, harmonization of norms and practices in the EU is still far from being achieved (Art29 2010b), and little is known about what factors incentivise, or prevent, the adoption of best data protection practices in the corporate world. An important gap seems also to exist between the way privacy is understood and disciplined in legal documents and the way privacy protection is translated into organisational practices, culture and procedures on the ground (Bamberger and Mulligan 2011). Demand for more robust regulatory measures to safeguard personal data does not only come from privacy advocates (PI 2009), but also from private companies, which think, for example, that organisations should be forced to disclose data breaches (Sophos 2010). In terms of solutions to tackle privacy and data protection problems, the debate is divided between proponents of legal solutions (Purtova 2009), and supporters of privacy-enhancing technologies (Weitzner, Abelson et al. 2008).

Understanding privacy in the era of big data poses new challenges (Tene 2012). Unfortunately no systematic study of the relationship between dimensions of value creation, through data analysis and brokerage, and the level of data protection has been carried out yet (Kshetri 2014). Previous studies suggest that companies which operate under strict regulatory privacy regimes tend to implement more protective internal privacy-preserving measures (Milberg, Smith et al. 2000). However employers and employees could make an exception to data protection rules in the presence of conflicts of interest such as the opportunity of using customers' personal data for making an extra sale (Ball 2010). A qualitative study based on semi-structured interviews with Chief Privacy Officers of 9 U.S. firms highlights the way the active role played by the US *Federal Trade Commission* in advancing a consumer-oriented understanding of privacy, as well as the passage of state security breach notification laws, have strongly contributed to make consumer privacy protection a market-reputation issue (Bamberger and Mulligan 2011). Further empirical investigation is needed to shed light on the corporate data protection environment.

1.3 Research questions, scope and implementation

This study represents an attempt to shed light on the tensions emerging from an organisation's desire to use personal data for fostering innovation and generating new services, and an organisation's concerns to avoid infringing privacy and data protection laws or to avoid violating people's privacy expectations.

Thus, this study proposes to investigate to what extent data protection principles are implemented inside enterprises, and how these decisions are influenced by two main factors: the degree of analytical sophistication an organisation has achieved and characteristics of the institutional context wherein the firm operates, namely the data protection regulatory regime.

Another important aspect, analysed in this research, concerns the relationship between the accessibility of analytical procedures and surveillance devices to gather and store personal data, and an organisation's privacy culture. The concept of privacy, and the legal regime it produces,

has been criticised by surveillance scholars for being unable to prevent massive data surveillance, or dataveillance (Clarke 1988, Gilliom 2011). For this reason, this research also explores the relationship between the current privacy and data protection legal regime and the deployment of dataveillance procedures within organisations.

Therefore, this study is built around two central research questions.

- **Research Question One:** How does the data protection regulatory regime influence enterprise data protection and data management decisions?
- **Research Question Two:** How does the level of analytical sophistication an organisation has achieved influence enterprise data protection and data management decisions?

Before moving into presenting the research design the terminology here adopted needs to be further clarified. The term enterprise is used, within this context, to indicate both for profit and nonprofit entities, as done in previous studies (Newman and Wallender Iii 1978). Thus, the term organisation is used as a synonym to indicate broadly both public agencies and private firms (Boyne 2002). Although entities operating in the not-for-profit and in the for profit sectors differ dramatically in the type of ownership, organisational structure, employees' job satisfaction (Yau-De, Chyan et al. 2012), and operating environment, it is possible to compare private and public organisations within functional categories such as information technology management features (Rainey and Chun 2005). From this perspective it has been shown, for instance, that high levels of "red tape" positively influence information technology innovativeness in entities operating in both sectors (Moon and Bretschneiber 2002). Organisational size and administrative capacity, for instance, influence process innovation in the case of both public entities (Walker 2014) and private firms (Cohen and Klepper 1996). As the dimensions here investigated—i.e. analytical sophistication and compliance with data protection principles—may play a role equally in for profit and nonprofit entities, the questionnaire used to gather the data was designed in such a way to allow people working in organisations operating in both sectors to participate in the study.

In order to gather information on the way enterprises manage conflicts of interest emerging from their willingness to collect, and analyse, vast amounts of data related to individuals, and the need to comply with data protection regulatory demands, an online survey was used. Between December 2013 and March 2014, business professionals working in areas related to privacy law and big data analytics (e.g. Chief Information Officers, data analysts, marketing directors, privacy professionals, and data protection experts) were invited to participate in the *Big Data Protection Study*. Information about the study and mechanisms to participate were advertised through professional groups' newsletters, online magazines, blogs, and specialised press. This strategy increased the chances to reach the target population of people working in the field of data protection law or in the area of big data analytics.

A website, *www.bigdataprotection.co.uk*, was also created by the researcher to offer further information about the study to potential participants. The electronic questionnaire was developed by the researcher through the Qualtrics online platform. In total 442 professionals accessed the electronic survey; 46% of them answered to all questions. The relationship between respondents and organisations was established by means of a set of screening questions which allowed the researcher to exclude retired or unemployed people. Detailed information on respondents' characteristics, limitations and generalizability of results are reported in section 6.2.1.

1.4 Contributions of the research

As data collection and processing practices have spread globally, the need for safeguarding personal information from unauthorised use has grown dramatically. Corporations all over the world have begun to appoint Chief Privacy Officers, as part of their information management strategies, to manage the risks of data breaches and poor compliance with national privacy laws (IAPP 2010). The theme of information privacy has gained momentum and it has been examined from different angles: from economics (Acquisti 2010), marketing (Lanier and Saini 2008) and

management information system (Il-Horn, Kai-Lung et al. 2007), to law (Solove 2006) and policy (Mulligan and Bamberger 2013), as well as public opinion polls (Margulis 2003).

Drawing on previous studies on business responses to privacy regulation (Milberg, Smith et al. 2000, Greenaway and Chan 2005, Khansa and Liginlal 2007, Ball 2010), this research explores the interplay between the degree of analytical sophistication and the level of compliance with EU data protection principles. It responds to the need to investigate privacy at organisational level (Culnan and Armstrong 1999), and to the need to explore the evolution of the corporate data privacy environment and to understand data management strategies (Bamberger and Mulligan 2011). Although this study mostly relies on privacy studies, it also draws insights from surveillance studies, particularly the concept of dataveillance (Clarke 1988, Degli Esposti 2014), in order to investigate the relationship between targeted analytics, an organisational privacy culture and the respect for data protection principles. By relying on the contribution of surveillance studies, this research hopes to shed light on the limits of addressing digital monitoring only from a privacy perspective (Stalder 2011, van Dijck 2014).

1.5 Structure of the thesis

The thesis is composed of seven chapters. The first three chapters introduce the reader to the topic of data protection and big data. They set the stage and present an overview of the academic literature, namely the interdisciplinary field of (a) privacy studies, with contributions coming from legal studies, information systems and marketing research; (b) surveillance studies; and (c) business studies exploring the adoption of big data within organisations. The other three chapters present the research framework, methodology, and the analysis and discussion of results.

Chapter Two and Three present the research context. While Chapter Two offers an overview of the debate around big data and analytics, Chapter Three describes the privacy and data protection European regulatory landscape. The concept of big data and key characteristics of analytically sophisticated organisations are reported in Chapter Two along with considerations

related to the evolution of the data economy and its implications for the safeguard of information privacy in the digital age. The discussion on information privacy continues in Chapter Three. This chapter pays attention to the European Data Protection regulatory regime, the privacy studies debate, and the relationship between data protection and information security from an organisational point of view. Benefits and limits of existing legal and technical solutions to protect information privacy are also discussed within this chapter.

While Chapters Two and Three rely on both legal documents, academic studies and grey literature, Chapter Four draws only on the academic literature to describe the specific knowledge gap in privacy studies this research is conceived to fill out; a set of propositions are also identified within this chapter. These propositions allow the researcher to build a detailed theoretical framework which is tested in Chapter Six. Thus, Chapter Four summarises insights gathered in the previous two chapters, presents the research question and builds the study's theoretical framework.

Chapter Five explains the methodological approach and the construction of the survey instrument, while providing detailed information on each construct, with definitions and corresponding survey items. Chapter Six is divided into two main sections. The first part of the chapter includes a presentation of the data, the data collection strategy, and the construction of the indicators used to measure all the constructs presented in Chapter Five. The second part of the chapter presents the statistical techniques used to test propositions identified in Chapter Four. Chapter Seven offers an extensive discussion of the limits and implications of the study's findings with conclusions and contribution to practice. Finally, the Appendix contains information about the project's website, the survey instrument, the study report produced for participants, and the blog posts published to invite professionals to contribute to the study.

CHAPTER TWO

The Big Data Promise

2.1 Introduction

Digital data are constantly created as a product, or by-product, of the many IT-based activities organisations and individuals perform every day. Blurring boundaries between the online and offline worlds and the growing digitisation of personal records have also made available an enormous volume of data about organisations' internal and external operations.

This relatively new trend, characterised by the rapid accumulation and processing of digital data in different formats, is known as 'big data'. By analysing big data organisations can radically improve their performance by becoming more efficient or by offering new services (Davenport and Harris 2007, McGuire, Manyika et al. 2012, Brynjolfsson and McAfee 2013, Davenport and Dyché 2013, Davenport 2014, Goes 2014, Wei-Hsiu and Woo-Tsong 2014). Furthermore, an entire digital economy is flourishing as a result of the production, collection, analysis and storage of digital data (Iyer and Davenport 2008, Naone 2008, Bernal 2010, Davenport, Mule et al. 2011).

Organisations take advantage and contribute to the data economy in several ways, from selling databases to analysing data for other organisations. However, as digital data are often linked to, or refer to, actual people, questions related to information privacy and organisations' data protection practices need to be addressed (Davenport, Harris et al. 2007, Tene and Polenetsky 2012). Big data technology brings risks alongside benefits. Practices such as massive data accumulation and analysis raise concerns on the potential negative consequences that the proliferation of monitoring devices and profiling tools may have on individuals and society (Clarke 1988, Lyon 1994, Marx 2006, Lyon 2007, Andrejevic 2009, Turow and Draper 2012, Wood and Ball 2013). This theme will be discussed at the end of this chapter and further explored in the next

chapter. Big data, analytics and information privacy are strictly intertwined phenomena: this chapter and the following one set the stage of this study by offering an overview of these topics.

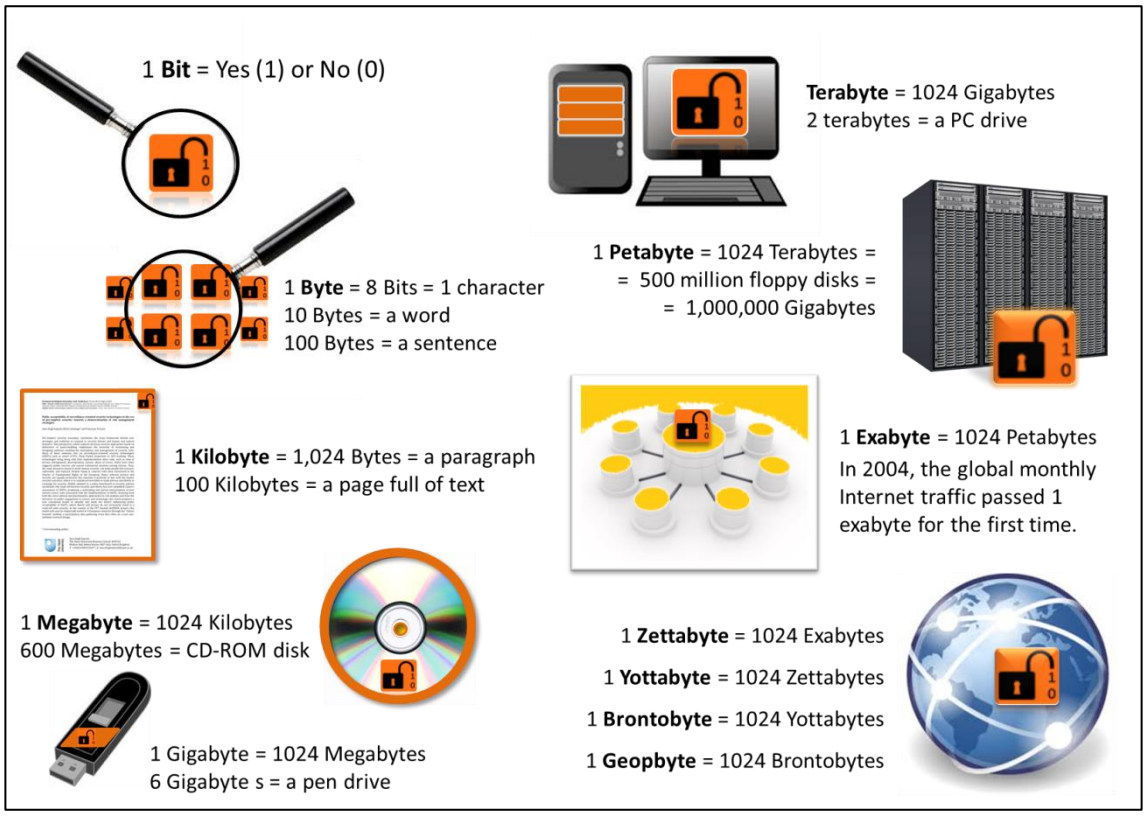
2.2 Defining Big data

The term 'big data' indicates *high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making* (Gartner 2013). For this reason big data is usually defined by three attributes, which are volume, velocity and variety, identified as the 3 'Vs' attributes. The term volume refers to the size of the data set, velocity indicates the speed of data in and out, and variety describes the range of data types and sources (Philip Chen and Zhang 2014).

In order to have an idea of what order of magnitude the attribute 'big' is meant to indicate, we may think of Google Inc., which currently processes over 20 petabytes a day of user-generated data (Scott and Bracetti 2013) in its eight data centres based in the US, Finland and Belgium. Figure 1 offers some hints to help us understand what a petabyte is. The proliferation of data gathering tools, such as sensors, and the retention of meta-data and other types of digital footprints contribute to increasing the size of the datasets at an exponential rate (Philip Chen and Zhang 2014).

Besides volume, big data is also characterised by data variety, which refers to the format of the data; namely, data can be structured (e.g., financial, electronic medical records, government statistics), semi-structured (e.g., text, tweets, emails), unstructured (e.g., audio and video), and real-time (e.g., network traces, generic monitoring logs) (Kambatla, Kollias et al. 2014). Velocity, which refers to the speed of the data acquisition and actualisation processes, contributes considerably to increase the complexity of big data problems.

Figure 1. Understanding units of measurements of digital information



Source: Author's elaboration of (Randy 2015).

Finally, *big data* is a term used not only to refer both to very large datasets holding a variety of data types, but also to the *analysis* of these data (Cavoukian, Stewart et al. 2014). Having data of good quality, stored in compatible formats, and easily accessible, represents fundamental preconditions to analyse information and get useful insights (Davenport, Harris et al. 2010). For this reason some commentators consider that 'veracity', meaning data quality, should be considered as another constitutive dimensions of big data (Dale 2015).

The fact that big data refer both to the platform used to manage vast repositories of data, and to the analysis of these data to create knowledge relevant for decision makers, causes sometimes some confusion. For this reason within this study we adopt the expression 'big data analytics' to refer to those technologies which help organisations leverage data to facilitate decision-making.

Big data analytics is data mining applied to very large datasets. As the size of the database grows, the information management infrastructure and the computational power necessary to apply advanced data mining tools change accordingly. The commercial application of analytics is of monetary value (Wei-Hsiu and Woo-Tsong 2014). Business process optimization and social-

network-based recommendations represent examples of big data applications with a considerable profit-enhancing potential (Kambatla, Kollias et al. 2014). The profitability aspect related to big data projects in the business context makes some commentators suggest that an additional 'V' attribute, namely 'economic value' (Hashem, Yaqoob et al. 2015), should also be added to the definition of big data.

Since this study investigates the way big data and analytics have been implemented across industries and within public and private organisations, the next section will explore how these terms relate to previous business nomenclatures, technologies and trends.

2.3 Origins of Big data

Initially, large international corporations have been the organisations more exposed to the challenges of managing high volumes of digital data. The fundamental difference between what is going on today and the past, is that in the 1970s data were internally produced and mainly referred to operational activities, with only a small portion of them coming from external sources (Watson, Wixom et al. 2006). In contrast nowadays data come from point-of-sale transactions, e-commerce or other online operations. The current almost negligible cost of storing data has also greatly contributed to the accumulation of historical data (Davey 2010).

Despite all the emphasis on the novelty of big data (Bernhut 2012), this phenomenon should be better interpreted as a form of incremental, rather than radical, innovation (Popadiuk and Choo 2006) prompted by three fundamental trends in IT (Laney 2001):

1) e-commerce generation of transactional data and organisational willingness to retain this information thanks to diminishing storage costs;
2) boosted data creation speed prompted by the more frequent interaction between organisations and customers both in store and online;
3) Availability of solutions for integrating and managing a wider variety of information, with different formats and structures.

From an information systems perspective it is worth noticing that, in the past, data was stored in data repositories usually owned, or managed, by the same company and was also customised to be processed by a few specific applications running on Decision Support Systems (DSSs); DSSs are IT systems that collect, organise and analyse data in order to facilitate and boost decision-making processes related to management, operations and planning within organisations. The path toward 'big data' was laid out in the late 1980s, when firms in the telecommunication, retail, and financial industries created their own data marts and data warehouses able to store vast amounts of customer and sales-related data, modelled in such a way to support a variety of applications (Watson, Wixom et al. 2006). Nowadays, in contrast, companies often rely on external service providers, specialised in storing, cleaning and securing the information of others, in order to protect, organise, and manage their employees' and customers' data. These companies offer not only data warehouse solutions, but also various business intelligence services, from database architecture to data modelling and analytics.

The difference between traditional DSSs, business intelligence applications and current big data applications can be found in the opportunities big data offer to apply sophisticated data mining tools, technically known as *knowledge discovery in databases* (KDD). Data mining is considered a mature technology which features "a necessary stress on algorithmic aspects and a preference for prediction" (Adams 2010: p. 18). Big data analytics allows running complex queries, and looking for correlations across the entire dataset. Examples of applications of big data analytics to business problems can be found in areas such as intelligent transportation systems, large-scale e-commerce, and Internet search indexing.

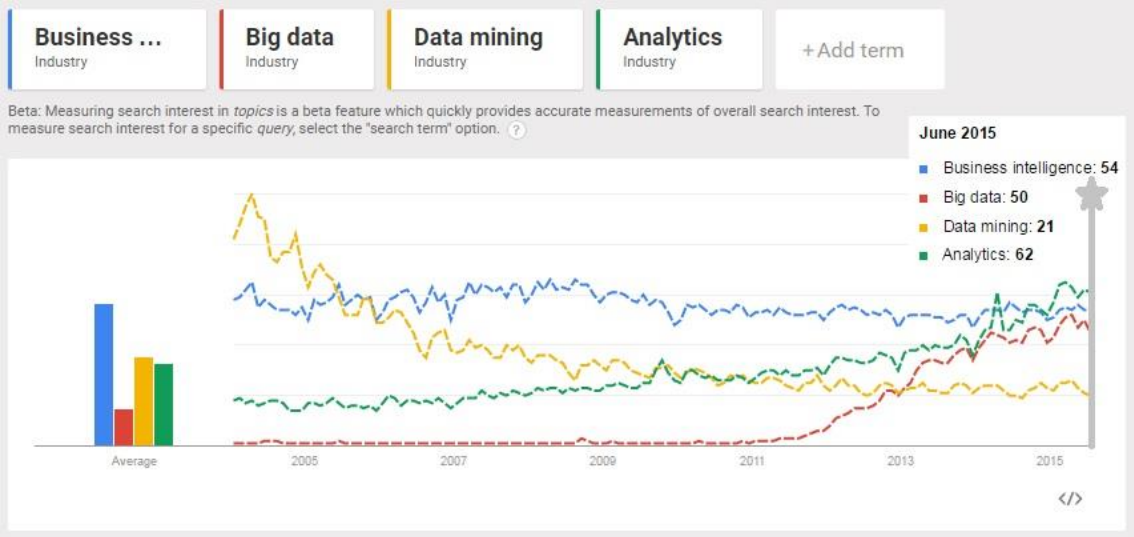
From a business perspective, analytics is where the real value and novelty of big data can be found. In the private sector, *big data analytics* is already becoming an integral part of several organisations' data management systems and business models because it enables: (a) experimentation to spot the source of variability in performance; (b) segmentation of populations to take customised actions; (c) computerisation of decisions through automated algorithms; and (d) the creation of new services, products, and business models (Manyika, Chui et al. 2011). The

ability of forecasting future business trends can also be enhanced by means of big data. Search data from search engines, for instance, have been proved to be fairly accurate estimators of future business activities such as housing market prices (Wu and Brynjolfsson 2012). By relying on data gathered from social media, text mining techniques have been successfully used to study web users' emotional reactions to the launch of new products or to the impact of catastrophic news (Sebor 2007).

To conclude, big data and analytics represent developments of already existing technologies. Big data analytics pursues the goal of improving performance by relying on fact-based evidence (Davenport 2014) exactly like *business intelligence* was used to improve business decision making through fact-based support (Negash and Gray 2008). To offer a visual representation of how the interest in big data, business intelligence, data mining and analytics has changed over time we can use Google Trend graphical tools as an aid. Google Trend is an example of how big data 'web' analytics can be applied to generate new services and collective knowledge.

Figure 2 shows how the interest in these key terms has changed between January 2004 and December 2014. The graph has been created by the author by using the online tool Google Trends. 'Interest' is measured in terms of the number of times a word has been searched on Google over a certain period of time. By looking at the chart, we can see how the interest in *business intelligence* has been almost steady over the past ten years (blue line on top), similarly for the interest in *analytics* (green line), which has benefitted in recent times from the attention given to *big data*, a term which appears in Google searches in the year 2011 (red line); in contrast, the term *data mining,* a very hot topic in the early 2000s, seems to have been replaced by *analytics* in recent times (crosswise yellow line).

Figure 2. Interest over time in the topics 'Business Intelligence', 'Big data', 'Data Mining', and 'Analytics'



Source: Google Trends (Google 2015)

Once the terminology has been clarified, the rest of the chapter presents an examination of organisational aspects related to big data as well as to the use made by private firms of big data analytics solutions. This brief overview will thus be followed by an examination of the way big data is used within organisations.

2.4 Big data and competitiveness

Technological progress plays a key role in boosting organisational competitiveness. By using data on information technology spending by 370 large firms, a study demonstrates that IT investments increase productivity and create substantial value for consumers (Hitt and Brynjolfsson 1996). The productivity associated with computerisation increased if we consider a time period of 5-7 years, which means that IT investments require a long-term vision to show their benefits (Brynjolfsson and Hitt 2003). Investments in information technology are capable of generating higher productivity both by reducing costs and by increasing service or product quality (Brynjolfsson and Hitt 2000). In further detail, computer-mediated transactions improve data extraction and analysis, controlled experimentation, personalisation and customisation (Varian 2010). Inventors and innovators also receive astonishingly high rewards and obtain control of the market in digital economies (Brynjolfsson, Malone et al. 1994).

Big data is one of the most recent technology trends. It is expected to benefit society and organisations in a number of ways. It has been estimated that, through the use of big data, Europe's public sector could potentially reduce the costs of administrative activities by 15 to 20 *percent*, through both efficiency gains and a reduction in the gap between actual and potential collection of tax revenue (Manyika, Chui et al. 2011). From the perspective of individual consumers and end-users big data brings several opportunities. For instance, the availability of price comparison websites may reduce costs for consumers and create new commercial channels: data aggregators, particularly infomediaries, accounted for about 33% of all motor insurance sales in the UK in 2012 (Breckenridge, Farquharson et al. 2014). The increased inventory carrying capacity of Internet retailers and consequent amplified product variety, made available through electronic markets, have also significantly enhanced consumer welfare (Brynjolfsson, Yu et al. 2003). Big data and the economy around digital data represent an opportunity to envision radical changes in the way many services are offered. Both the database and the analysis of the information contained in this database contribute to generate economic value. Since the cost of storing data becomes minimal, computing power increases, and more people perform their activities online, organisations have started realising the advantages of big data.

Big data has contributed to the reshaping of organisations inside and outside their boundaries. Digital technologies in the second machine age have disrupted traditional labour and capital (Brynjolfsson, McAfee et al. 2014). Investments in IT have already reduced the size of companies in terms of number of employees (Brynjolfsson, Malone et al. 1994); automation is increasingly substituting cheap labour. More specifically, sectors with high IT investments feature work systems characterised by decentralized authority, systems of incentives that compensate for decreased observability of decision makers' actions, and the predominance of knowledge workers (Hitt and Brynjolfsson 1997). These factors are stable across industries and are indifferent to variations in the measures used. The demand for more skilled labour is higher in firms which have invested in IT and have adopted complementary actions such as a decentralised workplace organisation and product/service innovation (Bresnahan, Brynjolfsson et al. 2002). Digital

enterprises are likely to gain predominance in the marketplace, especially when they are capable of becoming both market innovators and low cost producers (Keen and Williams 2013).

Digital firms which have heavily invested in proprietary technologies and network infrastructures, and have built platforms enabling key transactions are very likely to become market leaders. Google Inc. is a typical example of a company which has become predominant thanks to its ability to harness big data. Google not only monetizes consumers' intentions as revealed by their searches and other online behaviour, it also owns a scalable, efficient network infrastructure consisting of approximately one million computers, which run an operating system that allows new computer clusters to easily plug in (Iyer and Davenport 2008). It is so efficient that its estimated costs seem to be one-third that of its main competitors (Keen and Williams 2013). Data fusion and analysis can be considered strategic elements within organisations.

Highly successful companies characterised by innovative and sustainable business models are known as 'analytical competitors'. Analytical competitors are organisations that "use analytics extensively and systematically to outthink and outexecute the competition" (Davenport and Harris 2007: p. 23). Analytical competitors make use of advanced data mining techniques to transform data into 'actionable intelligence'. They are capable of transforming the knowledge extracted from the analysis of the data into profits and revenues. Actionable intelligence allows goal-directed knowledgeable actors to choose between the alternatives that have been presented to them as reasonable (Gandy 2012). Some commentators claim we have entered the big data 'Analytics 3.0' era (Davenport and Dyché 2013, Davenport 2014): an era when any type of firm will have the chance to become part of the new data-driven economy. We may say that success, measured in economic terms, demonstrates whether a firm has been able to apply analytics in the right way.

> It's important to remember that the primary value from big data comes not from the data in its raw form, but from the processing and analysis of it and the insights, products, and services that emerge from analysis. (Davenport and Dyché 2013: p. 30).

The computer and electronic products and information sectors, followed by finance, insurance, and government are the sectors that have gained most from the use of *big data*. However, not all companies operating online become market leaders: in fact great innovators are not immune from price or market erosion. Companies such as Amazon or Google are typical examples of highly successful and innovative digital businesses (Keen and Williams 2013). Other firms, such as Sony or Barnes & Noble represent cases of companies which have faced significant difficulties. Barnes & Noble was dependent on few domestic and international suppliers – a weakness which did not affect Amazon or Alibaba – and it has been involved in several legal proceedings leading to heavy expenses (MarketLine 2014). Sony Corporation saw on the 27th of April 2011 its Play Station Network (PSN) and other websites hacked by Luztec, which gained access to the personal information of more than 100 million users. Since then Sony has been a target of several cyberattacks perpetrated by groups opposing its copyright policy (Inagaki 2014). The company's involvement in various lawsuits and legal proceedings still adversely affects its brand image and stock price; it also puts an additional burden on the cost structure of the firm in terms of fines imposed and penalties levied (WMI 2014).

Big data can improve performance in different functional areas. The following critical areas are the ones which are considered to have benefited the most from the use of big data analytics (Manyika, Chui et al. 2011, Tankard 2012, Philip Chen and Zhang 2014).

1. Increasing operational efficiency
2. Informing strategic direction
3. Improving customer service
4. Developing new product and services
5. Enhancing customer experience
6. Identifying new market
7. Entering new markets

8. Complying with regulation.

This list is neither complete nor exhaustive. Big data analytics has also been successfully used to improve systems' and users' security. In the field of fraud detection, for instance, data analytics has been a key driver for many years now in identifying what constitutes normal and abnormal patterns of activity (Constantine 2014).

Firms' internal marketing units have a long tradition in mining data to investigate consumer preferences and behaviour. By analysing customers' past transactions, attributes and online behaviour, marketers can tailor products and offer in greater detail. Personalisation and customisation represent a particularly rapidly evolving and innovative business realm (Davenport and Harris 2007, Davenport and Harris 2007). Analytics can be used to pursue several marketing objectives (Davenport, Harris et al. 2010, Davenport and Dyché 2013, Davenport 2014), such as: customer selection, loyalty, and service, i.e. identification of customers with the greatest profit potential, customer retention and loyalty; pricing, i.e. identification of the price that will maximize yield or profit; product and service quality, i.e. detection and minimization of quality problems.

In addition, data analytics can be used to achieve several other objectives in a variety of functional areas (Davenport and Harris 2007, Davenport 2014), such as: supply chain management, i.e. simulation and optimization of supply chain flows, reduction of inventory and stock-outs*;* human capital, i.e. selection of the best employees for particular tasks or jobs, and compensation levels; financial performance, i.e. analysis of financial performance and effects of nonfinancial factors; research and development, i.e. improvement of quality, efficacy, and safety of products and services*.*

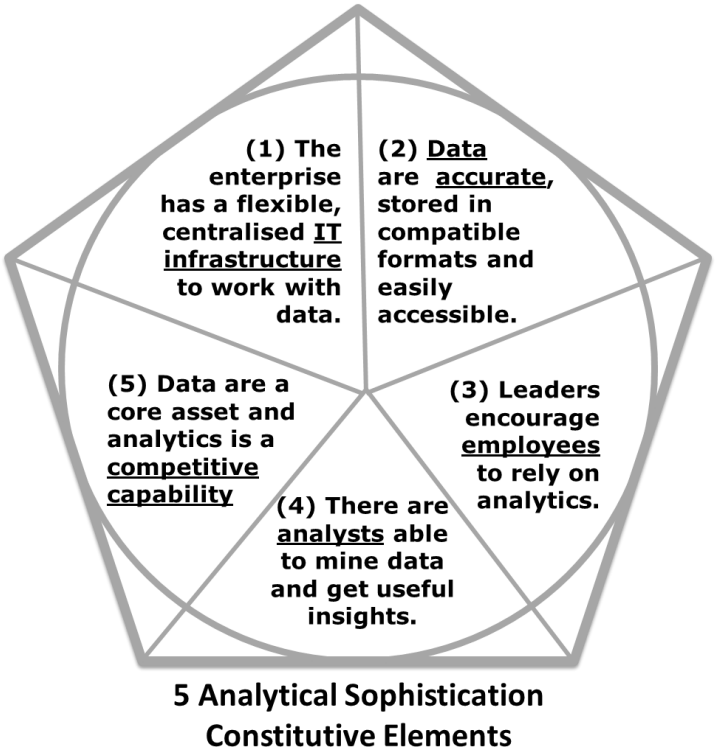
To better understand how companies can become analytically sophisticated, in the following section we will explore in greater detail those specific characteristics which differentiate analytical competitors from their peers.

2.5 Analytical competitors' characteristics

Five elements characterise analytically sophisticated companies. These elements, which are summarised by the acronym DELTA (data; enterprise; leaders; targets; analysts) (Davenport, Harris et al. 2010), describe a specific mix of organisational and technological components, which are listed in figure 3 and can be summarised as follow.

- Analytically sophisticated organisations keep their **data** accurate and easily accessible.
- In doing so, the **enterprise** owns an integrated and flexible infrastructure to access and work with data.
- Firm **leaders** understand the importance of analytics and encourage employees to make use of it.
- Data analytics represents a distinctive, competitive capability of the organisation, developed to achieve specific **targets**.
- The organisation employs **analysts**, who are people with the necessary statistical and programming skills to mine data and get useful insights from it.

Figure 3. Analytical competitors' DELTA features



Source: Author's elaboration of (Davenport, Harris et al. 2010).

Companies which produce large amounts of data as part of their operations are the ones more likely to go through the path to become an analytical competitor. For example, firms in the financial sectors, including securities, investment services and banking, have the highest data-intensity per firm on average (Manyika, Chui et al. 2011). In other words, the financial sector stores, on average, the largest amount of digital data and financial service providers heavily invest in IT both to process and protect their data. In order to keep data complete and up to date it is necessary to invest in data warehouse and integration. An integrated information system platform ensures that data are constantly updated, cleaned and accessible to people from different units within the organisation. The information management system allows the firm to control information on several aspects of the enterprise and to focus on those functions which are critical for achieving strategic objectives. As explained in the following quote, investing in the IT system is a necessary condition for becoming an analytical competitor (Davenport and Harris 2010).

> You need to make a commitment to conceiving of data as a competitive advantage. The next step is to build out a low-cost, reliable infrastructure for data collection and storage for whichever line of business you perceive to be most critical to your company. If you don't have that digital asset, then you're not even going to be able to play the game. And then you can start layering on the complex analytics. Most companies go wrong when they start with the complex analytics. – Statement made by Jeff Hammerbacher, Chief Scientist at Cloudera (Rappa 2001).

Another challenge to become an analytical competitor is represented by leadership. For a company to be considered an analytical competitor, business leaders should be able to accept and include the insights produced by analysts into the decision-making process; they should believe in the value of taking decisions based on empirical evidence and not use statistics as a way to improve the legitimacy of decisions which have already been made (Brydon and Gemino 2008, Brynjolfsson, Hammerbacher et al. 2011). For this reason, firms able to compete on analytics usually are managed by 'IT savvy' CEOs (Haggerty 2012), who are keen on taking 'fact-based decisions'. Supporting timely strategic decision-making is the main reason behind investing in business intelligence applications and big data projects (Davey 2010).

Besides improving decision-making, analytics usually contributes to 'build a distinctive competitive capability' (Davenport, Harris et al. 2001) in one or more business functions, such as marketing, operations, information security or finance, within an organisation or a business group. Data analysts and people with the 'necessary mathematical and statistical skills' should also be employed by the organisation. The proliferation of the analytical culture among business is also pushing the demand for analytical talents: for example, in the US several data and analytics job opportunities are already offered on dedicated web sites like *iCrunchData.com*.

Organisations are also drawing insights directly from science to improve their business processes (Davenport 2009). Business experiments, which apply the scientific method to determine whether a particular business intervention is effective or not, are increasingly used to improve

products and services (Davenport 2007). By 'scientific method' we refer to the systematic observation and measurement, as well as the formulation, testing, and modification of hypotheses through experiments. Experiments test the causal relationship linking two observations, as predicted by some theory, while controlling for other factors. By running experiments organisations can assess the effectiveness of actions already implemented (Davenport and Harris 2009). They are, for instance, well suited to assess the efficacy of attributes of marketing product and promotions, as well as webpage design, economic incentive schemes, and other easy-to-measure business initiatives (Davenport 2007, Davenport 2009).

Thus, big data is also creating a new space of encounter for academia and the private sector. Increasingly psychologists and other social scientists see collaborations with online businesses and Web 2.0 platforms as an opportunity to run large scale experiments and test concurrent hypotheses at limited cost. Similarly, digital enterprises have also begun recruiting social scientists asking them for help in analysing the huge amount of data about users' attitudes and behaviours. Users of online services are often unaware of the fact they are participating in an experiment. The deployment of big data analytical solution to influence people's behaviour has considerable privacy implications. These and other aspects related to privacy and big data will be explored in the next section on dataveillance.

2.6 Big data, privacy and dataveillance

Collection, availability, and migration of data are fundamental preconditions to any kind of use of them a company may envision. It would be impossible to tackle the problem of protecting data from abuse, without understanding the forces behind the same creation, assemblage and processing of data. To accomplish this task we need to draw insights from a different stream of literature in order to understand the desire to continuously reconstruct digital footprints into profiles. Rather than focusing on the idea of privacy this area of inquiry adopts, as a key fundamental idea, the concept of surveillance.

Surveillance refers to the monitoring and supervision of populations for specific purposes (Lyon 2001). Dataveillance refers to "the systematic use of *personal data systems* in the investigation or monitoring of the actions or communications of one or more persons" (Clarke 1988: p. 499). It is more covert than any form of traditional physical surveillance, because it is applied not to the individual themselves, but to a data-shadow of a real person. New surveillance technologies are augmenting the power of the 'surveillant Other' through the creation of a 'surveillant assemblage' that easily scrutinises and targets people's digital identities or 'data doubles' (Haggerty and Ericson 2000). Dataveillance, conceived as a complex set of procedures and techniques used to collect and organise data about individuals, sustains the feeling of oppression of traditional surveillance, but adds to it fears of the unseen and unknown, and significant risks of error, ambiguity and misinterpretation (Clarke 1994). "Social control is the element that most fear with regard to computerized surveillance, and thus it features—alongside privacy—most prominently in discussions of new technology" (Lyon and Zureik 1996: p. 3). Furthermore, technological changes, such as the proliferation of radio-frequency identification (RFID) sensors or ubiquitous computing, have contributed to facilitate and magnify surveillance by making it relatively inexpensive (Bankston and Soltani 2014).

Surveillance studies interpret dataveillance as a way to exercise and intensify surveillance practices and processes, already present in modern societies, through database management (Lyon 1994). According to this line of inquiry dataveillance might participate, as a powerful 'social sorting' mechanism, to the enactment of contemporary modes of social reproduction (Lyon 2002) and social exclusion (Amoore and DeGoede 2005). It may also pose dangers to the individual, such as arbitrariness, behavioural manipulation (Degli Esposti 2014), discrimination, or unjustified exclusion or persecution, as well as to society, by enacting a prevailing climate of suspicion and fostering adversarial relationships (Clarke 1994). Privacy and data protection measures tend to be limited to individualistic readings of the situation, and not to consider issues of fairness and equality (Lyon 2001).

However, surveillance has two faces: advantages appear alongside serious disadvantages (Lyon 2001). Mass dataveillance can help foresee the epidemic of a disease or improve understanding of collective actions and communications. Provided that the normative framework allows for an appropriate use of personal data for the sake of public good (Fortin and Knoppers 2009), evidences coming from different fields show how the increasing availability of data about human attributes and behaviours may foster basic and applied scientific research (McCarthy 2000, Wood 2000, Ayres 2007, Weston, Hand et al. 2008). It has also been argued that the proliferation of information about individuals' involvement in the criminal justice system, financial distress, or other embarrassing activities, could reduce distasteful statistical discrimination caused by the lack of transparency and the recourse to stereotypes (Strahilevitz 2008).

Even though dataveillance could create positive outcomes, it is certainly risky. Personal data can be stolen, manipulated, misinterpreted, sold, and disclosed to the detriment of both data-subjects and data-processors. In trying to single out suspicious behaviour, vulnerable groups can suffer from controversial interpretations of what constitute abnormal behaviour (Levi and Wall 2004, Amoore and DeGoede 2005). Big data analytics can also influence people's autonomy and self-determination. Data accumulation and analysis can be interpreted as ways to obtain social control by means of digital surveillance (Gandy 1993, Lyon 1993, Lyon 1994, Ball and Wilson 2000, Lyon 2001).

The use of big data analytics and an organisation's internal data to perform marketing actions or other types of interventions is not the only controversial aspect of big data from a privacy perspective. A growing market for digital data has emerged alongside the market for big data products. While the big data industry features vendors of technology for storing, transmitting, and analysing large quantities of dynamic and diversified structured or unstructured data for social or commercial purposes, the data economy includes a large variety of data brokers, infomediaries, search engines and other enterprises whose business model heavily relies on the collection and exploitation of data for commercial purposes. These data, which are often personal data, are merged with other data, sold and processed to obtain different objectives.

Firms which process large amount of data as part of their operations can also benefit from sharing or selling them to other firms (Acquisti 2010). Customer databases represent a new category of proprietary asset; their capacity to produce income contributes to increase overall firm value. In 2008, for instance, the publishing company *Reed Elsevier*, currently called *RELX Group plc*, agreed to buy *ChoicePoint Inc.* for $3.5 billion (Thiel 2008) in order to acquire C.L.U.E. (Comprehensive Loss Underwriting Exchange), a database created by *ChoicePoint* containing up to seven years of information provided by insurance companies about personal property claims history.

Because of the importance of these players in shaping the big data landscape, the next section presents an overview of their evolution and characteristics.

2.7 The data economy

The growth in data availability has given rise to a number of business types, which form the data economy. These business types are presented in this section; a summary of their characteristics has been included in table 1.

Data aggregators gather information from disparate sites and package it for users, either in the Web community or on intranets (Taylor 1996): the information comes from multiple sources and itis all neatly organised to be understandable at a glance. Many famous data aggregators originally offered, like libraries, offline consultation services. For instance, Reed Elsevier Lexis-Nexis, a widely used database that stores laws and court acts, as well as news and magazine articles, was initially conceived as an offline legal research system for Ohio statutes, created from the initiative of The Ohio State Bar Association in 1973 as Lexis (Plosker 2004). At the beginning data aggregators were developed for an offline world, though their expansion almost thirty years ago helped the development of internet protocols and infrastructures, and the enlargement of internet-based services has subsequently driven the emergence of a new type of data aggregators known as *search engines*.

Search Engines are an important part of the data economy. Part of search engines' economic value consists of information about service users. Search engines such as Google, Yahoo!, Microsoft MSN Search—active as Bing—, and AOL account for almost the total number of search queries conducted globally. Google alone cover more than half of the total market. Information about web-site users is sold to advertising companies for targeting purposes. This information combined with clickstreams allow marketers to improve the effectiveness of their online marketing campaigns and to reduce costs by paying only for those web users who have actually visited the advertised company web site.

Data Brokers are companies which collect and sell customer data to other companies. Certain data brokers specialise in background checking and rate people according to specific characteristics such as financial solvency. Their main areas of activity are credit scores and background checks for employment, insurance and loan purposes. Yet commercial data brokers also sell dossiers to the government for law enforcement purposes (Hoofnagle 2004). The US Federal Trade Commission (FTC) defines a broker of customer data in the following terms:

> [d]ata brokers are companies that collect information, including personal information about consumers, from a wide variety of sources for the purpose of reselling such information to their customers for various purposes, including verifying an individual's identity, differentiating records, marketing products, and preventing financial fraud. (FTC 2012: p. 68).

Craig Spiegelberg, CEO and founder of *Location Sentry*, LLC, says that the data broker business in the United States generates $150 billion a year. "This is done by a host of unknown companies that, like spyware, check what data you have uploaded when, places you shop, or destinations you travel. The sensors collect data, including GPS locations, texts, call recording, contact lists, what data is sent to what IP addresses, and a lot more.." (Grundvig 2014). In the collection and sale of consumer data specifically for marketing purposes, data brokers seem to operate with minimal transparency (FTC 2012, US-Senate 2013).

> Data brokers typically amass data without direct interaction with consumers, and a number of the queried brokers perpetuate this secrecy by contractually limiting customers from disclosing their data sources. Three of the largest companies – Acxiom, Experian, and Epsilon – to date have been similarly secretive with the Committee with respect to their practices, refusing to identify the specific sources of their data or the customers who purchase it (US-Senate 2013: p. ii-iii).

Despite important differences in the legal environment, data sellers and brokers are present in Europe as well. In Spain, for example, not only are company reports on *eInforma.com* and SMEs contacts details on *laGuía.es* on sale, but also information about private citizens is easy to obtain. For about ten Euros anyone can purchase information about someone's home and work addresses, employment and marital status, names of relatives and neighbourhoods, cases of law violation, and more on *Dateas.com*. According to the Spanish data protection law, privacy is not under threat if the petitioner can demonstrate the legitimacy of the request (Dateas 2010). Family reunion is an example of a legitimate request.

Data brokers establish alliances with other businesses to exchange customer data or purchase or license data from retailers, financial institutions or other data brokers. Government and census data, and any sort of publicly available data, such as professional certifications or hunting or pilot licenses, are also important consumer data 'avenues' (US-Senate 2013). By means of 'warranty cards, sweepstakes entries, and other types of surveys' customers also voluntarily share they data to enter a prize competition, or to earn money, shopping vouchers or coupons. The UK-based marketing firm *Caci* runs a geo-demographic tool called *Acorn*. On the base of its detailed life-style classification of all 1.9 million British postcodes, Acorn helps to determine where to locate business activities and to identify public health and educational performance needs. The data broker *Experian*, besides providing real time profiling services through its *Mosaic* classifications, offers also personal credit-score assessments.

In some cases data brokers are called infomediaries. Formed from a combination of the words 'information' and 'intermediary', an infomediary is a Web site that gathers and organises large amounts of data and acts as an intermediary between those who want the information and those who supply the information. The nature of the information traded or shared also contributes to the specialisation of data brokers in determined niches. Credit rating agencies are special *rating agencies*, which are a type of data broker which provides risk assessment evaluations of institutions or individual citizens.

Finally, companies specialised in analysing data for other companies are named by the press as *data crunchers* or *data miners*; these are consultancies specialized in extracting useful patterns for action from huge amounts of personal data. These companies can offer insights on how to understand different kind of behavioural patterns, from voting during political elections to grocery purchases (Baker 2009, Flavelle 2010). Data crunchers, which are firm specialised in data mining, can emerge from a composite set of job experience across marketing, informatics and analytics, as a response to data availability. This was the case of the British firm *Dunnhumby*, which was founded by Clive Humby and his wife Edwina Dunn in 1989, and then acquired by its major client, Tesco (Wood and Lyons 2010).

In general, digital enterprises (Rappa 2009) consider users' data as a key asset and use them as a new type of currency in managing their transactions and negotiations with other organisations. Revenues in the digital world mainly come from online advertisement (Anderson 2009). Google's search-based advertising, for instance, represents a way to monetize consumers' intentions as revealed by their searches and other online behaviour (Iyer and Davenport 2008). Despite its long-term mission – approximately 300 years according to CEO Eric Smidt – is "to organise the world's information and make it universally accessible and useful" (Google 2015), currently Google's income mainly comes from search-based advertising.

New business models have emerged to monetise the value of people's data. Any business model is essentially a set of key decisions that collectively determine how a business earns its revenue,

incurs its costs, and manages its risks (Girotra and Netessine 2014). Ensuring that these new business models are compatible with users' information privacy expectations is not only a challenge, but an imperative for businesses that want to avoid customers' disillusionment and distrust toward online services and e-commerce (Featherman, Miyazaki et al. 2010).

Table 1. Summary of typical data firms

Type of firm	Definition	Example
Data Aggregators	*Data aggregation* is any process in which information is gathered and expressed in a summary form, for purposes such as statistical analysis. A common aggregation purpose is to get more information about particular groups based on specific variables such as age, profession, or income. The information about such groups can then be used for Web site personalization to choose content and advertising likely to appeal to an individual belonging to one or more groups for which data has been collected (Rouse 2014).	*Reed Elsevier* is a provider of professional information solutions in the Science, Medical, Risk, Legal and Business sectors. It combines personal data sourced from multiple public and private databases. *ChoicePoint*, now wholly-owned by Reed Elsevier, maintains more than 17 billion records of individuals and businesses, which it sells to an estimated 100,000 clients in 2005, including 7,000 federal, state and local law enforcement agencies (ChoicePoint 2009).
Search Engines	A Web search engine produces a list of "pages"—computer files listed on the Web—that contain the terms in a query. Most search engines allow the user to join terms with and or, or to refine queries. They may also let users search specifically for images, videos, news articles or for names of Web sites (Britannica 2014).	Google Inc., American search engine company, founded in 1998 by Sergey Brin and Larry Page in Mountain View, California, is a search engine firm which now offers more than 50 Internet services and products. More than 70 percent of worldwide online search requests are handled by Google, which in 2011 Google earned 97 percent of its revenue through advertising based on users' search requests. (Britannica 2014)
Data Broker	An individual who or organisation that searches for information for clients. Information brokers, also called information retrieval consultants or resellers, use various resources including the Internet, online services that specialize in databases, public libraries, books and CD-ROMs, or plain old fashioned telephone calls, to build accurate users dossiers.	Acxiom collects detailed information about people and provides a 13-digit code for every person who is placed into one of 70 lifestyle clusters, ranging from "Rolling Stones" to "Timeless Elders" (Behar 2004). Acxiom customers include nine of the country's top ten credit-card issuers, as well as nearly all the major retail banks, insurers, and automakers.

Infomediary	The term was originally used to define a company that works as a personal agent on behalf of consumers to help them take control over information gathered about them for use by marketers and advertisers (Hagel and Rayport 1997).	*Autobytel* offers consumers a place to gather information about cars and car companies before they make purchasing decisions. Autobytel acts like a messaging services: buyers submit their information, a message goes to the interested dealers, and then the dealers get in contact with the potential buyer.
Data Cruncher	To transform raw data into useful insights, programmers have to recycle legacy data, translate from one vendor's proprietary format into another's, check configuration files, and yank data out of web server logs. This kind of programming is usually called *data crunching*, a synonymy of data mining, which is the process of analysing data from different perspectives and summarising it into useful information.	Tesco's Dunnhumby was an example of a company offering analytical services to retailers and brands and helps companies in develop their Customer Relationship Management strategies.

2.8 Conclusions

This chapter has provided an overview of the big data phenomenon. Big data and analytics are complex, multi-layered realities which are exercising their influence on firms, markets and society in a number of ways. Some firms, such as analytical competitors, data firms and big data vendors, are especially benefitting from the availability of digital data and the proliferation of advanced data mining tools. The next chapter will explore big data phenomenon from a different angle by presenting data protection and security aspects related to big data processing. Considerations about the implications of big data for people's information privacy and autonomy will be also discussed, as well as firms' considerations regarding compliance with privacy laws and motivations behind investing in information security.

CHAPTER THREE

Data Protection Laws and Organisations' Data Privacy Strategies

3.1 Introduction

Among the achievements of the 20th century in terms of freedom and fundamental rights there is the concept of privacy. Article 8 of the *European Convention on Human Rights* (ECHR), signed in 1950, states the duty to respect one's "private and family life, his home and his correspondence."

Privacy became a central topic of the academic computer-science debate in the United States in the 70s, when the proliferation of data banks, and computer mainframe systems with remote access, enabled previously unconceivable and highly pernicious governmental and business data-gathering activities (Gürses and Berendt 2010). Considerations related to the effect of being under surveillance on human behaviour and the risks it poses to individual autonomy (Askin 1972), motivated law-makers' discussions on the need for limiting the potential chilling effects of subtle IT surveillance practices (Miller 1972). Memories of the pervasiveness and danger of totalitarian regimes informed this need of imposing boundaries to any type of activity that might lead to mass manipulation (Flaherty 1989).

As explained in the previous chapter, massive data collection, and analysis, represents nowadays not only a fundamental pillar of the digital economy, but also a source of privacy concern. This chapter presents a specific aspect of the European institutional environment: the data protection regulatory framework, its basic principles and its potential effects on organisational information management and security strategies. By drawing insights from legal and business studies literature on the matter, this chapter investigates the potential impact of regulation, especially privacy law, on the internal organisational privacy culture, and how this would affect other

organisational choices related to relying on analytics as a source of competitive advantage. As said in the last part of the previous chapter, which was devoted to the concept of *dataveillance*, the final aim is to study the relationship between data protection and data usage in order to identify both synergies and points of friction.

This chapter offers an overview of the legal framework governing privacy in the EU and of the main streams of research which have investigated privacy and data protection within business studies. The research question will be presented at the beginning of the next chapter, followed by a critical review of the two streams of literature presented in the second and third chapters in search of theoretical insights to shed light on the issue at stake. The first part of this chapter summarises the evolution of privacy laws in developed countries. The second part of this chapter presents the central idea of information privacy concerns and interesting findings of the limited number of empirical studies that address this topic. It also highlights those areas of limited knowledge where further investigation is required.

3.2 Overview of the legal privacy landscape

Over the past 30 years privacy has moved from being a third-level social policy issue to become a first-level social and political issue in all advanced economies. During the 70s several European legislatures adopted statutes concerning the processing of personal data – examples are the 1973 Swedish Data Act, and the 1977 West German Data Protection Act. In 1980 the Council of Europe signed Convention 108 to reassure every individual, whatever his/her nationality or residence, about their right to privacy, with regard to automatic processing of personal data relating to them. In 1970 the *Fair Credit Reporting Act 'FCRA'* (FTC 2012) was issued to govern consumer reporting agencies in the United States; a context characterised by rapid technological innovation and widespread consumers' privacy concerns (Westin 2003). It was followed, in 1974, by the *Privacy Act*, (USC 1974), conceived to regulate the way federal agencies maintain records about individuals.

Recent history was an influential factor in shaping Europeans' views on privacy as well. In Germany the existence of extensive repositories of personal data, gathered by both public and private entities, enabled the Nazi regime to identify minority groups, and not only efficiently seize their assets, but also to persecute them (Flaherty 1989, Samuelson 2000). The anecdote is still used to explain why Europeans are particularly sensitive about privacy (Whitman 2004). In 1970 the German state of Hesse was the first to promulgate a data privacy law, which aimed at ensuring the functioning of a democratic society by limiting the potential threats arising from the increasingly sophisticated processing of personal data.

Cultural and historical differences informing the public's beliefs and perceptions lead European legislators to recognize privacy as a human right *per se*, without any need to derive it from other liberties (CoE 1981, EC 1995). In the Treaty of Lisbon it was granted the status of a fundamental right (EU 2007). A summary of the main European legislative acts is reported in table 2.

Table 2. Evolution of European Data Protection Legislation

European Legislation	**Reference**
Directive amending Directive 2002/22/EC, Directive 2002/58/EC, and Regulation (EC) No 2006/2004	(EC/136 2009)
Treaty of Lisbon amending the Treaty on European Union and the Treaty establishing the European Community	(EU 2007)
Directive amending Directive 2002/58/EC	(EC/24 2006)
Directive 2002/58/EC on privacy and electronic communications	(EC/58 2002)
Directive 2002/22/EC on universal service and users' rights	(EC/22 2002)
Directive 2002/21/EC on a common regulatory framework for electronic communications networks and services	(EC/21 2002)
Regulation No 45/2001 on the protection of individuals with regard to the processing of personal data by the Community institutions and bodies and on the free movement of such data	(EC-R-45 2001)
Directive 2000/31/EC on electronic commerce	(EC/31 2000)
Charter Of Fundamental Rights Of The European Union	(EU 2000)
Directive 99/5, 95/46 complemented by 2002/58, recommendation 1997/18	(EC 1999)
Directive 1997/66/EC on the processing of personal data and the protection of privacy in the telecommunications sector	(EC/66 1997)
Directive 1995/46/EC on the protection of individuals with regard to the processing of personal data	(EC/46 1995)

European Legislation	Reference
Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, known as Convention 108	(CoE 1981)
OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data	(OECD 1980)

In 1995, the European Commission approved *Directive EC/46 on the protection of individuals with regard to the processing of personal data, and on the free movement of such data* (EC/46 1995). The Data Protection Directive is an overarching regulatory framework, whose objective is the preservation of privacy as a fundamental human right and its protection from abuse. It established that special safeguards must be granted to data which refer to 'identifiable individuals'. This type of data refers to 'data-subjects' and are handled by 'data controllers' or 'processors'. Any organisation, private or public operating within EEA, which is collecting and processing *personal data*—i.e. data about identifiable persons—has the obligation to comply with data protection principles and restrictions. An explanation of the terminology adopted in the Directive is reported in table 3.

Table 3. EU data protection terminology

Term	Meaning
Personal data	Personal data means any information relating to an identified or identifiable natural person or 'Data Subject'.
Identifiable person	An identifiable person is someone who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his or her physical, physiological, mental, economic, cultural or social identity.
Sensitive data	The processing of special categories of data, defined as personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and of data concerning health or sex life, is prohibited, subject to certain exceptions (see Article 10 of Regulation (45/2001).
Data Controller	The Data Controller means the Community institution or body, the Directorate-General, the unit or any other organisational entity which alone or jointly with others determines the purposes and means of the processing of personal data. For each processing operation, a Data Controller must be identified and prior notice must be given to the Data Protection Officer of the institution.
Data Subject	The Data Subject is the person whose personal data are collected, held or processed by the Data Controller.
Data Processor	If the Data Controller does not execute the processing of personal data himself, this processing operation is carried out by a Processor on behalf of the Controller. He has to provide sufficient guarantees in respect of the technical and organisational security measures required and ensuring compliance with those measures. The Processor can be a natural or legal person, public authority, agency or any other body, acting on

	instruction, and only on instruction, from the Controller. Controller and Processor need to be bound by a contract or legal act for the carrying out of the processing operations of personal data.

Source: Author elaboration of "Chapter 2: Data protection terminology" (EU-FRA and CoU 2014).

The last requirement forced countries outside Europe to revise their privacy legislation or to agree about special rules to avoid stopping data flows. The Directive required European member states to update their legislations accordingly, to institute national enforcement agencies, and to allow the transfer of personal data only to countries able to guarantee an adequate level of protection. In 2000, after two years of consultations, EU and US signed an agreement, called International *Safe Harbor Privacy Principles*: a voluntary certification program for US companies ensuring the adoption of data protection principles and the safe transfer of data (US-EU 2000). Nonetheless, legal experts consider that there is a far lower standard of legal protection of online privacy in the US than in the EU (Baumer, Earp et al. 2004).

The reality of the digital economy, characterised by entities operating in different jurisdictions, poses considerable challenges to the ambitions of national privacy laws to protect citizens' data globally. Radical differences exist even between historical allies such as the EU and the US; for instance, in the US, privacy is a property right whereas in the EU, it is a fundamental right. Since July 2013, revelations about the US National Security Agency' surveillance practices, exposed by whistle-blower Edward Snowden, have further increased the divide between US and EU approaches to data protection (Aaronson and Maxim 2013). However, the reality of transatlantic data demands that both the United States and the European Union need to accommodate their privacy cultures by revising programmes such as Safe Harbor (Colonna 2014).

Following different social, cultural and legislative traditions, several countries around the world have adopted privacy regulatory regimes (See table 4). In general countries have opted for enforcing either (a) *sector regulation*s, like the US, (b) *comprehensive/omnibus regulations*, like the EU, or (c) *self-/co-regulatory approaches*, like for example Australia (Bellman, Johnson et al. 2004). Principle-based approaches have been developed in Europe, India and other Asian

countries (Bajaj 2012) as well as in South and North America. However, while in the EU the principle-based legislation is virtually valid in any circumstance, privacy legislation in the US follows a case-by-case logic.

Table 4. Correspondence between Western data protection normative frameworks

	OECD Guidelines	Fair Information Practice Principles (FIPPs)	Directive 95/46/EC
OBJECT	Personal data, whether in the public or private sectors	Personal information managed through information practices	'Personal data' i.e. any information relating to an identified or identifiable natural person ('data subject').
AIM	To provide a minimum standard	To provide adequate privacy protection	To protect natural persons' right to privacy with respect to the processing of personal data, without neither restricting nor prohibiting the free flow of personal data.

Source: Author's elaboration of (HEW 1973, OECD 1980, EC 1995, FTC 2000).

In Europe, data protection law features some basic data protection principles. Since these principles represent a common understanding on the way privacy rights should be protected in practice, the next section presents these principles in further detail.

3.3 Basic data protection principles

Privacy and data protection laws and guidelines tackle the problem of protecting data from abuse through a set of principles that, when respected, let organisations process personal data in a fair and lawful way. As showed in table 5, similar data protection principles constitute the backbone of privacy regulations across all Western countries showing a substantial policy convergence (Bennett 1992). The comparison between four sets of principles, that is, the 1973 United States Fair Information Practice Principles (HEW 1973), the 1980 OECD Guidelines on the free flow of information, the 1995 EU Data Protection Directive, and the 2004 United Nation principles of good practice (UN 2004), demonstrates the similarity. Although it is easy to identify the same core ideas in each set of principles, the emphasis assigned to each principle, which can be differently reworded, may change in a significant way across jurisdictions.

Table 5. Correspondence between data protection principles across jurisdictions

	World	OECD countries	US	EU
	United Nations	***OECD Guidelines***	***Fair Information Practice Principles***	***Directive 95/46/EC***
	(UN 2004)	*(OECD 1980)*	*(HEW 1973, FTC 2000)*	*(EC 1995)*
1.	Collection Proportionality	Collection Limitation Openness	Notice/Awareness	Notice Disclosure
2.	Transparency Use Transfers/Disclosure	Purpose Specification Use Limitation	Choice/Consent	Purpose Consent
3.	Quality Access and Correction Objection	Individual Participation	Access/Participation	Access
4.	Security	Data Quality Security Safeguards	Integrity/Security	Security
5.	Accountability	Accountability	Enforcement/Redress	Accountability

Source: Author's elaboration of (HEW 1973, OECD 1980, EC 1995, FTC 2000, UN 2004).

Since the '70s, *fair information practice* became the dominant US approach to information-privacy protection (HEW 1973) and sectorial privacy regulations the prevailing regulatory modality (Westin 2003). The code of Fair Information Practice Principles (FIPPs) was conceived to increase government transparency with high aspirations (Reidenberg 1994). In fact, according to the 'Code of Fair Information Practice' there must be no personal data record keeping systems whose very existence is secret; there must be a way for an individual to find out what information about him/her is in a record and how it is used and to prevent information about him/her that was obtained for one purpose from being used or made available for other purposes without their consent.

Despite the existence of these principles, a comprehensive approach to privacy has not materialised in the United States, maybe due to the lack of consensus and clarity about the nature of the interest that individuals may have in their personal information, and legislation was created on a piecemeal basis (Hoofnagle 2010). A wide array of statutes—both at federal and at state level—governs the collection, use, and dissemination of personal information. Furthermore, business groups tend also to frame privacy in terms of securing personal information. This approach may lead to undermine other aspects of fair information practices, like limitation of data collection, with respect to security (Hoofnagle 2010).

The EU Data Protection Directive 1995 includes similar data protection principles, which can be divided into those which deal with data controllers' obligations (see table 6) and those which refer to data subjects' rights (see table 7). As showed in table 6, data controllers and processors operating in Europe must comply with six principles to ensure they manage personal data in a lawful way. They have to keep data accurate, updated and free of error. Data have also to be deleted after a certain time. Data should be collected for specific purposes and not used outside the scope for which they were collected. Data must also be protected from unauthorised use and people must be made accountable for data mishandling. Finally, data transfer should follow strict rules and be managed with caution.

As reported in table 6, five core principles can be identified. These are: notice; consent; access; security; redress. The first principle requires organisations to inform consumers about the data collection purpose/s, identity of data gathering agencies, and data-retention period. Besides being informed, according to the second principle, individuals have also the right to authorise the organisation processing their data to collect, retain, use and transfer data to third parties. The third principle expresses the possibility for people to access the data in order to correct errors, or to demand the deletion of their data. According to the fourth principle, data holders have to protect data from any form of data breaches that could result from intentional actions, such as hacking, employee theft, theft of physical equipment, as well as negligence or the accidental loss of laptop computers or other hardware, unintentional exposure on the Internet, or improper disposal of data. Finally, data holders must be accountable for any damage caused through data loss or unauthorised disclosure.

Table 6. EU Data controllers' obligations

Obligation	Meaning
DATA QUALITY	Personal data that are collected and stored should be accurate and reviewed periodically to ensure that they are kept accurate and up to date.
PURPOSE SPECIFICATION	The collection of personal data should be limited to data that are adequate and relevant for the specified purpose or purposes.
RETENTION	Data destruction procedures may be as important for the protection of an individual's privacy interests as the process of data collection and retention.
DATA TRANSFER	There should be no disclosure, or transfer, or other use except those needed to achieve the purposes specified when the data were collected. Personal data should not be transferred to third parties unless the individual was informed that such disclosure may take place and provided that it can be ensured that the data will be given the same level of protection by the recipient as was provided by the sender.
DATA SECURITY	Appropriate security measures should be implemented to protect against risks presented by the collection, use and storage of an individual's personal data, whether from accidental loss, damage or disclosure or deliberate interference.
ACCOUNTABILITY	Data controller compliance should be ensured through a system of enforcement, which includes sanctions for those who handle data inappropriately.
REDRESS	The system of compliance should ensure the ability of a data subject to seek redress for breach of the principles in the processing of his or her personal data.

Table 7. EU Data subjects' rights

Right	Meaning
RIGHT OF NOTICE	Collection of personal data should be done fairly and lawfully. Fair collection means that an individual should be informed, at the moment of collection, of the contemplated uses of that data and of the purpose(s) for collecting data; who will be using the data, who is in charge of protecting those data, and, if applicable, any contemplated transfers of the data and to whom.
RIGHT TO CONSENT	Personal data should not be disclosed, made available or otherwise used for purposes other than those initially specified except with the consent of the data subject.
RIGHT OF ACCESS	Individuals should have the right to inquire whether their personal data are being used and the right to obtain a copy of all personal data collected and maintained that relate to them.
RIGHT TO OBJECT	Individuals should have the right to object to the processing of the personal data relating to them in certain situations, such as where serious damage or distress results.

The effectiveness of data protection principles in the area of digitisation, service personalisation, and ubiquitous surveillance has been criticised by a number of commentators. Empirical studies have demonstrated the limits of applying a contractual and consent-based approach to the problem of users' participation in data handling decisions (Sheehan 2005, Hoofnagle, Soltani et al. 2012), and legal and business commentators have pointed out the challenges of complying

with the 'purpose specification' principle in the context of big data projects (Mayer-Schönberger 2010, Cate, Cullen et al. 2013, Mayer-Schonberger and Cukier 2013). Studies on the limits of applying de-anonymisation techniques in current times characterised by the availability of large amounts of personal information posted on social media (Acquisti and Gross 2009) have also contributed to generate a contentious debate between the supporters and the opponents of a right and principle-based approach to the protection of individual privacy (Cavoukian and Castro 2014).

Yet all commentators agree on the drawbacks of limiting privacy talks to mere regulatory solutions and understand that technological, organisational, psychological, sociological and economic aspects have also to be taken into consideration at the time of tackling the issue of information privacy. For this reason, the next section offers an overview of the multi-disciplinary privacy studies field and opens the discussion on the alternative ways and motives driving data privacy and information security decisions.

3.4 Privacy studies: an overview

The problem of privacy has been tackled in ethics and law, economics and business studies, psychology and sociology. On one side, economic theory has studied the data-subject/data-holder dyad and their respective privacy/transparency trade-offs (Acquisti 2010). On the other side, social psychologists have put considerable effort into assessing privacy concerns across different groups and over time (Margulis 2003). Legal scholars have been tracking the evolution of digital and data mining technologies to limit and prevent discrimination and the violation of privacy (Zarsky 2002). The privacy battle field has mainly been the public arena. Several scholars have studied the determinants and the consequences of individual reactions toward the use of intrusive digital technologies across different cultural settings (Zureik, Harling Stalker et al. 2010).

Neoclassical economics starts from the assumption that market agents will always disclose favourable information and withhold negative information about goods quality. According to this

perspective, any kind of legislation, such as privacy legislation, which creates opportunities for concealing information from the market, would thus create market distortions and inefficiencies. As a result, privacy legislation might have negative implications for markets (Posner 1978). This perspective used to have some explanatory power in contexts wherein information was a scarce resource. In the digital era, however, characterized by information overproduction, economists, especially behavioural economists, have started recognizing the value of privacy for both consumers and markets. Namely, the modern microeconomic theory of privacy considers privacy regulation necessary to safeguard individual and collective welfare (Acquisti 2010).

Besides economics, privacy has also been studied in psychology. Scholars have analysed psychological states and functions of privacy which apply at personal and group level. Most of the theories formulated out of these studies have been quite influential in informing our general understanding of privacy. Altman considers privacy a dynamic process of interpersonal boundary control and defines it as "the selective control of access to the self" (Altman 1975). Similarly, Westin emphasizes the importance of self-disclosure, intimacy, anonymity and reserve in the safeguard of one's mental integrity and in the development of relationships (Westin 1967). Although privacy is considered a cultural universal, social, environmental, cultural and socio-developmental factors may change its manifestations quite significantly (Margulis 2003). Although the same concept of privacy has been criticized for being hyper-individualistic and for hiding the central issue of discrimination underlying any manifestations of surveillance (Gilliom 2011), it not only represents a set of enduring policy instruments, but it may serve as a tool for privacy advocates to resist the excessive monitoring of human behaviour (Bennett 2011).

> In the emerging information economy, privacy will not primarily mean preventing organisations and other people from knowing about us. Instead, it will be founded on securing organisations' commitment to principles about what shall and, crucially, shall not be done with those data. Privacy cannot be an absolute right, but will remain a centrally important value. Privacy can best be understood as a protection against certain kinds of risks – risks of injustice through such things as unfair inference, risks of loss of

> control over personal information, and risks of indignity through exposure and embarrassment (Six 1998: p. 2).

Because of the interdependence between individual privacy expectations, participation in online data-sharing activities, and the impact of customers' reactions to corporate and governmental privacy-invasive initiatives, a large literature has dealt with the topic of privacy in business studies. Although privacy is considered to be an elusive, evolving and culturally embedded social construct (Ribak 2007), *information privacy,* defined as the ability of the individual to personally control information about one's self (Stone, Gardner et al. 1983), is a concept widely used and accepted in the marketing and information management literature (Bélanger and Crossler 2011). Because of the importance of these studies in understanding privacy at the crossroad between users and organisations, the next section will focus specifically on the study of information privacy from a business studies perspective.

3.5 The business studies' viewpoint: studying information privacy

Information privacy refers to the power of data-subjects to choose whether to reveal, or conceal, information about themselves to others. In economics and computer science information privacy has widely been framed in terms of 'control over'—or 'access to'—private information as the disclosure of sensitive information could carry a potential for vulnerability that must be monitored (Pavlou 2011). Information privacy is a fundamental part of information management and it has been widely discussed in the information systems literature with attempts to go beyond technological aspects in order to capture wider societal implications (McFarlan 1988).

Within the context of this study the idea of information privacy is more useful than the idea of general privacy which combines privacy of personal communications with privacy of data (Clarke 2006). Besides this, the notion of 'control over information' is an underlying assumption of European data protection legislation, which gives data subjects the right to have a say in the way

their data are used, shared and kept, with important implications for public self-determination and empowerment (Whitley 2009).

Organisational and information management scholars have largely studied privacy at an individual level; namely they have paid attention to antecedents and consequences of information privacy concerns (Il-Horn, Kai-Lung et al. 2007, Son and Kim 2008). The primary dimensions of individuals' concerns about organisational information privacy practices – which refer to concerns about data collection, errors, secondary use of information and improper access – have been measured through a reliable scale in the context of both offline and online transactions (Smith, Milberg et al. 1996, Malhotra, Sung et al. 2004). As showed in table 8, there is an interesting correspondence between the dimensions of individual privacy concerns and global data protection ideas.

Table 8. Correspondence between 'privacy concerns' dimensions and data protection principles

Dimensions of Privacy Concerns*	**Corresponding Data Protection Principles**
Collection	(1) Notice/Disclosure
Unauthorised Secondary Use	(2) Consent/Purpose
Improper Access	(4) Security (5) Accountability
Errors	(3) Access

Source: Author's elaboration of (Smith, Milberg et al. 1996)

Individual privacy attitudes have also been investigated in several marketing studies. While marketing studies about privacy expectations usually try to explain customers' dissatisfactions with services or communication campaigns (Phelps, Nowak et al. 2000), information management studies tend to investigate how threats to privacy influence end-users' willingness to engage with online activities (Paine, Reips et al. 2007) and information systems end-user evaluations of security and privacy dimensions (Hui, Teo et al. 2007, Son and Kim 2008).

Thus, several studies have dealt with modelling and understanding why, how and to what extent people are concerned about their privacy. It has been empirically demonstrated that the more a person feels vulnerable to adverse consequences of information disclosure, the more concerned about her privacy she will be (Dinev and Hart 2004). In contrast, more experienced internet users

tend to be less concerned about their privacy (Bellman, Johnson et al. 2004). But, persons who have experienced previous privacy invasions begin withholding information (Hui, Teo et al. 2007) and are less willing to be profiled for personalized advertising (Awad and Krishnan 2006). Statements devoted to ensure online users about the safeguarding of their privacy have a positive effect on the propensity to provide information (Hui, Teo et al. 2007).

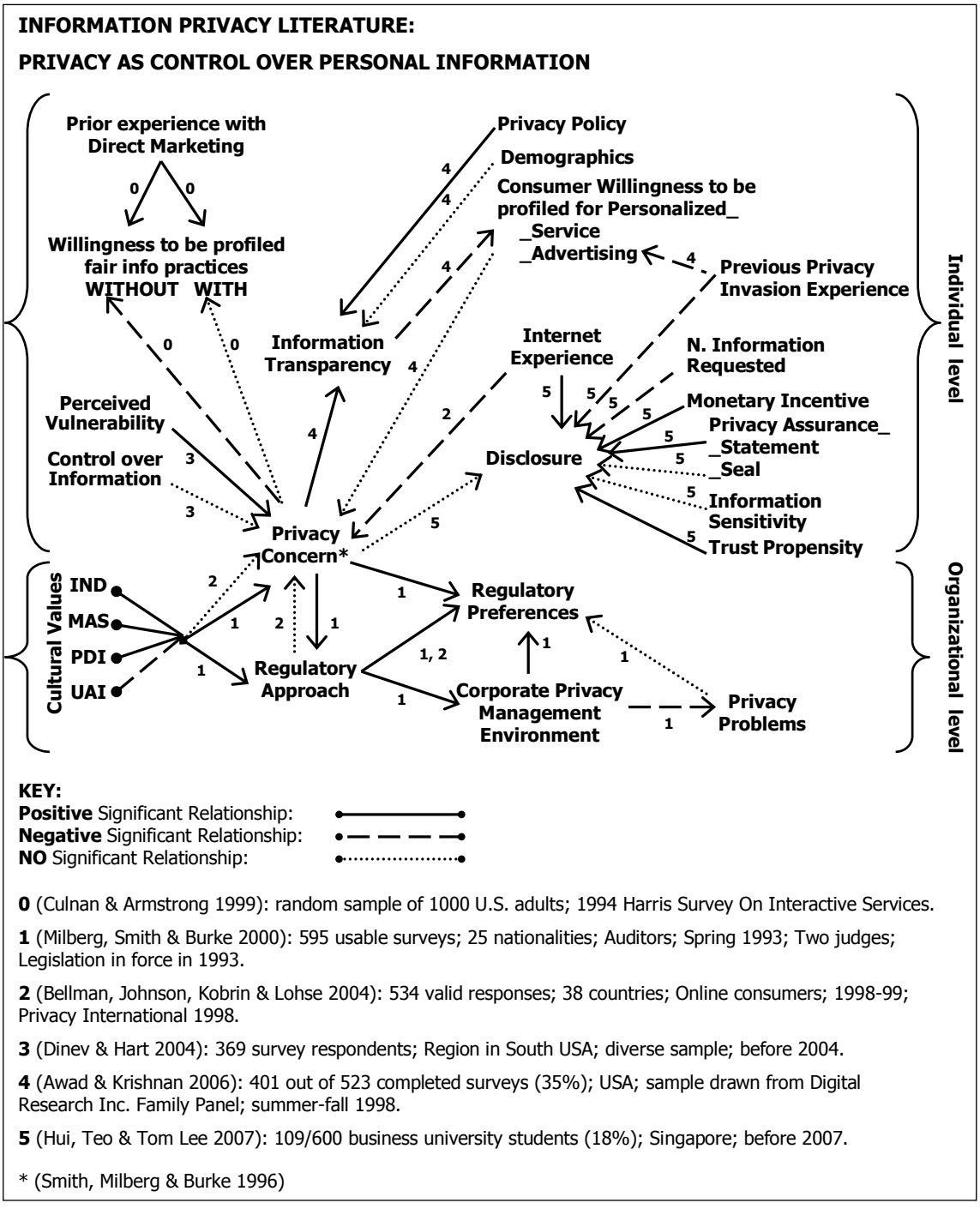
Psychologists and behavioural economists have largely contributed to the study of those factors that affect individual privacy evaluations (Altman 1975, Margulis 2003, John, Acquisti et al. 2011, Brandimarte, Acquisti et al. 2012). An aspect which has captured researchers' attention is the so-called privacy paradox. The privacy paradox refers to the tendency of online users to disclose a variety of personal details on social media sites while declaring to be concerned about their privacy (Barnes 2006). This sort of counterintuitive behaviour has captured the attention of psychologists (Trepte and Reinecke 2011), behavioural economists (John, Acquisti et al. 2011), and researchers working in the field of human-computer interaction (Kehr, Kowatsch et al. 2015).

So, people seem to surrender personal information for a small incentive but turn secretive when they suspect they are being observed (Hardin 2015). A synthesis of research conducted in the area comes to the conclusion that the complexity of understanding privacy from a user's perspective is due to the fact that privacy concerns are malleable and can be manipulated by governmental or private agents, concerns are also context-dependent and there is uncertainty related to the consequences of engaging in privacy-risky behaviours (Acquisti, Brandimarte et al. 2015). Furthermore, the identification of the so-called privacy paradox can also be produced by methodological decisions and support for this interpretation, for instance, has not been found when the theory of planned behaviour has been applied (Dienlin and Trepte 2015).

Although these studies have generously contributed to the comprehension of those factors that are influenced by, or influence, individuals' concerns about organisational information privacy practices (Smith, Milberg et al. 1996), privacy studies have mainly devoted attention to the study

of individual privacy preferences, with only a few studies investigating privacy at organisational level (see figure 4).

Figure 4. Map of relationships studied in the Information Privacy Literature



As the present study contributes to this second stream of research, the next section will offer recognition of previous studies which have investigated the topic of privacy from an organisational perspective.

3.6 Organisational privacy studies

Several factors can lead an organisation to believe privacy is important.

When information privacy becomes a shared value within an organisation, and employees adopt privacy-protective practices and procedures, an information protection culture is likely to emerge. An *information protection culture* can be defined as:

> a culture in which the protection of information and upholding of privacy are part of the way things are done in an organisation. It is a culture in which employees illustrate attitudes, assumptions, beliefs, values and knowledge that contribute to the protection and privacy of information when processing it at any point in time in the information life cycle, resulting in ethical and compliant behaviour (Da Veiga and Martins 2015: p. 249).

The expression 'information life cycle' refers to the various phases the information goes through from collection and processing until deletion.

Organisations can enhance their privacy programs by creating a culture of integrity that combines a concern for the law with an emphasis on managerial responsibility for the organisation's organisational privacy behaviours (Culnan and Williams 2009). Thus, the construct 'organisational privacy culture' captures the extent to which privacy is seen to be part of the company's culture (Greenaway and Chan 2013). Organisational or corporate culture is expressed in the collective values, norms and knowledge of organisations, which affect the behaviour of employees. An organisational or corporate culture is made of the pattern of basic assumptions, attitudes and beliefs of employees (Schein 1985).

Consumers' privacy preferences, or concerns, matter to business, especially to those operating in the business-to-consumer market. Consumers seem to penalise companies that do not adopt fair information practices while dealing with their data. In fact, the more concerned consumers are about their privacy, after controlling for the level of customers' familiarity with direct marketing initiatives, the less they would be willing to be profiled for targeted marketing if the company

does not adopt fair information practices (Culnan and Armstrong 1999). In contrast, when companies adopt fair information practices, privacy concerns have no effect on someone's willingness to be profiled for targeted marketing. But, interestingly, consumers who value more information transparency features, like data removal and data time-expiration, are the ones less likely to participate in personalization/profiling initiatives (Awad and Krishnan 2006).

The formulation of the legislation quite often also responds to mass media coverage of privacy scandals, data leaks and security breaches. Public concerns often represent the benchmark for setting privacy norms. Namely, opinion-pool research has been widely used to understand public privacy concerns in order to inform privacy policies (Gandy 2003). Cultural values and the level of information privacy concern influence the regulatory regime adopted (Milberg, Burke et al. 1995).

In a study which addresses the relationship between privacy concerns and organisational features, privacy professionals and auditors coming from different organisations offer insights about how companies approach the issue of data protection (Milberg, Smith et al. 2000). This study confirms that national privacy regulations reflect cultural values and public privacy concerns. Moreover, companies operating in jurisdictions characterized by less permissive privacy regulations are more likely to adopt safer internal practices, and to experience less privacy problems. Another very interesting result of this study is that, in presence of strong privacy concerns, restrictive privacy regulation in force, and a safe corporate privacy environment, the demand for comprehensive privacy laws increases. In other words, it seems to observe a self-reinforcing trend toward more protective privacy measures resulting from the interplay between inter-organisational privacy policies and extra-organisational privacy regulation.

Another stream of research has studied organisational privacy practices by providing assessments or metrics of the degree of compliance of privacy policies with codes like Fair Information Practices (FIP) principles (Ryker, Lafleur et al. 2002, Peslak 2005, Sheehan 2005, Schwaig, Kane et al. 2006, Storey, Kane et al. 2009). Studies have examined the extent to which FIP are present in the privacy policies posted on the websites of the business-to-business and business-to-consumer

websites of high technology US firms (Ryker, Latteur et al. 2002); and the most heavily trafficked and popular sites on the Internet (Culnan 1999, Culnan 1999). These studies come to the conclusion that, although internet privacy policies describe an organisation's practices on data collection, use, and disclosure, they do not address actual people's concerns but they tend to be used to comply with the law.

Privacy policies are demonstrated to be of limited use in increasing public awareness and participation in the management of digital data. They are difficult to understand, which causes people do not read them (Vail, Earp et al. 2008). The average length of the privacy policy of the top 50 Fortune companies is 1,581 words (Peslak 2005). The complexity of this policies would require users to have a postgraduate degree in law (Baumer, Poindexter et al. 2004). Besides lack of readability (Milne, Culnan et al. 2006), policies rarely contain information to enable the exercise of people's data access rights or any security assurance (Sheehan 2005). Privacy policies are mostly used to notify people their data are collected and ask for their consent (Peslak 2005, Sheehan 2005, Vail, Earp et al. 2008). A study which looked at privacy policies from nearly 50 websites and surveyed over 1000 Internet users revealed a notable discrepancy between what privacy policies are currently stating and what users deem most significant (Earp, Antón et al. 2005). Policies are also not easy to comprehend, and, as a result, online consumers frequently do not read them (Vail, Earp et al. 2008). To make meaningful choices with regard to their financial and medical Personally Identifiable Information (PII), consumers would have to devote significant amounts of time to studying the options available (Baumer, Poindexter et al. 2004) in very complex privacy policies (Schwaig, Kane et al. 2006).

Furthermore, several websites, even religious church websites, are at risks of adopting poor data privacy procedures (Hoy and Phelps 2003). The issue of information security deserves further attention as it constitutes one central data protection requirement, especially when it refers to database security management (Lipton 2001, Spears and Barki 2010). While organisational decisions related to information privacy are mainly driven by regulatory compliance and users' preferences, investments in information security are determined by a wider set of factors.

Organisations manage a large variety of types of proprietary information, from business plans or patents to clients' information. Protecting this information from malicious attacks or accidental disclosure plays an increasingly important role within modern organisations. For this reason the next two sections are devoted to exploring the issue of organisational information security decisions.

3.7 Cyber security threats

Modern society's overall dependence upon information technology and communication systems brings new threats alongside considerable benefits (Furnell and Warren 1999). Examples of cybercrime – broadly defined as a crime that employs a computer network during any phase (Kshetri 2006: p. 33) – include: online fraud, online money laundering, ID theft, use of computers to further traditional crimes, and cyber extortions.

The ever-evolving cyber threat landscape features financially-motivated cyber-criminal activity such as unauthorised access, online extortion and Distributed Denial of Service (DDoS) attacks, but also Jihad-oriented sites designed to facilitate radicalisation among the Muslim community, or the creation of counterfeit cards by organised crime groups by using Personal Identification Numbers (PINs) stolen from malware or skimming devices installed at ATM or POS terminals (Choo 2011). In phishing attacks, for instance, phishers send emails that mislead their victims into revealing credential information such as account numbers, passwords, or other personal information to the phisher (Hamid and Abawajy 2014). As most phishing emails are nearly identical to the normal emails, it is quite difficult for the average users to distinguish phishing emails from non-phishing ones.

Police forces in most countries face the challenges of dealing with the ubiquitous global nature of cybercrime (Kshetri 2006, Kshetri 2013). Cyber warfare is a topic of global concern and several States have developed National Cyber Security Strategies and Programmes (Robinson, Jones et al. 2015). Countries like UK, Germany, France, The Netherlands, and also US, Canada, Australia,

Russia, Estonia and Japan, have recognised in their national cyber security strategies the need to develop, or enhance, 'situational awareness', and national critical information infrastructures' resilience in order to maintain an open yet secure cyberspace (Franke and Brynielsson 2014). The cyber space is increasingly considered a new military domain.

> There is another domain […] similar to the seas in its sheer magnitude, seeming ubiquity, and lethal potential, but it is also unique in that it is not comprised of water and waves; rather, it consists of zeros and ones, optic fibers and photons, routers and browsers, satellites and servers. This is, of course, the […] Cyber Sea (Stavridis and Parker Iii 2012).

The complex and multi-layered nature of the cyberspace has generated massive criminal parasitism (Maillart and Sornette 2010). The convergence of wireless and wired IP networks (Samani 2007), the expansion of the smart energy grid (Knapp and Langill 2015), or business practices like the Bring-Your-Own-Device (BYOD) tendency (Kurpjuhn 2015), make it increasingly complicated to protect organisations' information systems and operations. The interplay between physical and digital elements in modern cyber security attacks complicates the issue further.

Vulnerabilities of the physical and the digital perimeters of the organisation are strictly intertwined and can be exploited for a number of different reasons, from theft of property rights to foster unfair competition, to military espionage or cyber hacktivism. China, for instance, is considered to be responsible for many cyber espionage operations which arguably begun in 2003 with a series of intrusions of US government and contractor networks collectively referred to by the code name Titan Rain (Lindsay, Cheung et al. 2015). The Chinese People's Liberation Army is engaged in 'information confrontation' and is pursuing a highly ambitious cyberwarfare agenda that aims to link all service branches via a common platform capable of being accessed at multiple levels of command (Inkster 2015). The director-general of the UK Security Service (MI5) in 2008 took the unprecedented step of writing a letter to three hundreds chief executives and security advisers of private-sector corporations alerting them to the threat of cyber exploitation from

China. The letter highlighted the director-general's concerns about the possible damage to UK business resulting from electronic attacks sponsored by Chinese state organisations, and the fact that the attacks are designed to defeat best practice IT security systems (Borland 2008).

Although breaches of sensitive personal information occur frequently, and under widely varying circumstances, the lack of reliable information makes it impossible to assess their actual volume and many incidents go daily undetected (GAO 2007). Data processing may represent a marginal activity for several companies, in spite of the amount and sensitivity of the type of information managed in their operations. Relying on external providers can be seen by organisations as a solution to cope with a firm's insufficient motivation and competence in protecting data. However, the practice of externalising data processing brings new risks and a lack of control over information.

> The most pervasive concern among CISOs may be the need to protect data that resides throughout an increasingly porous network, while expending precious resources on compliance. Compliance alone is not equal to being secure—it is simply a minimum baseline focusing on the needs of a special regulated environment. Security, meanwhile, is an all-encompassing approach that covers all business activities (CISCO 2014: p. 18).

As much of the information we have on cyber-crime losses is derived from surveys, it is difficult to give an accurate estimation of the costs and magnitude of cybercrime (Florencio and Herley 2011) and make a case within the organisation for the importance of investing in information security. Furthermore, the more companies invest in cybersecurity, the more cyber incidents they will be able to detect (PWC 2014). As explained in the next section, different elements play a role in determining an organisations' level of investments in information security and privacy safeguards.

3.8 Rationale behind information security investment decisions

In 2007, 827 professionals participated in a Deloitte & Touche's and Ponemon Institute's study on data protection and security risks. 85% of survey respondents said that they had suffered some kind of incident compromising personally identifiable information (PII) within the previous year (Deloitte 2007). Most firms realise the need to revise their information management practices after suffering a serious security incident. TJX, a retailer of apparel and home fashions in the United States, used to collect and store all customer information in order to either authorise purchases through credit card payment or accept returned items without a receipt. In December 2006 the retailer suffered a serious electronic security breach, which generated several fraud incidents. As a result of this event the firm undertook a deep revision of the firm's information management and security procedures (Tirial 2009).

The relationship between information security investments and vulnerabilities was initially established in an economic model developed by Gordon and Loeb. The model determined the optimal amount of resources to invest in information security to achieve various information security goals, such as protecting the confidentiality, availability, authenticity, non-repudiation, and integrity of information (Gordon and Loeb 2002). The model suggested that little or no information security investment is economically justified for extremely high, as well as extremely low, levels of vulnerability and corresponding economic losses. Furthermore, to maximize the expected benefit from investment to protect information, the optimal amount to spend on information security should never exceed 37% of the expected loss resulting from a security breach. In a subsequent study, Tanaka and co-authors verify the relation between vulnerability and information security investment by identifying a relationship between the existence of information security policies and vulnerability levels in Japanese municipal governments (Tanaka, Matsuura et al. 2005).

Since the year 2001 the interest in research on economic aspects of information security has increased dramatically (Gordon and Loeb 2006). Information security investment (Gordon and

Loeb 2002, Campbell, Gordon et al. 2003, Cavusoglu, Mishra et al. 2004) represents one of the central topic discussed in the information security studies literature (Wang 2012); other topics are: information security management and assessment (Eloff and von Solms 2000, Dutta and McCrohan 2002, Kotulic and Clark 2004); information security techniques (Refregier and Javidi 1995, Tajahuerce and Javidi 2000); information systems security monitoring and development (Hoffer and Straub 1989, Baskerville 1993, Straub and Welke 1998); and cryptographic technology design (Wang, Yin et al. 2005).

Economic theory can help explore how regulation, market dynamics and economic incentives can prevent or encourage individuals and organisations to invest in information security (Anderson 2001). The payment of high litigation costs and regulatory sanctions can motivate organisations to invest in information security. Bruce Schnieier suggests that if software vendors were liable for security vulnerabilities in their products, they would invest more in secure software development; and that, if liabilities were transferable among firms, they would start demanding cyber-insurances, which would push security vendors to demonstrate the ability of their products to reduce cyber risks (Schnieier 2002). In countries where data breach notification is compulsory, it has been demonstrated that individuals who suffer the loss or theft of their personal information are more willing to file a lawsuit against the company responsible for the incident, especially if the event has produced economic harm (Romanosky, Hoffman et al. 2014). Furthermore, firms in highly-regulated sectors seem invest more in cybersecurity than firms in unregulated sectors (Chai, Kim et al. 2011).

> This year, banking and finance respondents spent as much as $2,500 per employee (median) on cybersecurity, while retail and consumer products businesses invested up to $400 per employee (median) and education respondents invested a maximum of $200 per employee (median). […] Only 38% of survey respondents said they have a methodology to prioritize security investments based on greatest risk and impact to the organisation's business strategy (PWC 2014: p. 12-13).

The sensitivity of the data processed by an organisation contributes also to determine the level of security a company wants to guarantee to the treatment of its information (Gantz and Reinsel 2010). The 1995 Data Protection Directive requires data controllers and processors to implement information security measures necessary to safeguard the integrity of the data. Data classification policies help organisations allocate resources and treat data differently according to their relevancy and sensitivity. Thus, data can be considered: (a) just *private* (e.g. email address on a YouTube upload), or (b) necessary to comply with other regulations (e.g. emails that might be discoverable in litigation or be subject to retention rules); *custodial* (e.g. account information, a breach of which could lead to or aid in identity theft); *confidential* (e.g. trade secrets, customer lists, or confidential memos); and *lockdown* (e.g. financial transactions, personnel files, medical records, or military intelligence).

Cyber incidents and data breaches can also translate into a tangible decrease in a company's market value (Acquisti, Friedman et al. 2006) produced by reputational damage, a breakdown of consumer trust, or the loss of valuable intellectual property to a competitor (Lee, Kauffman et al. 2011). Different models have been developed to help enterprises allocate resources to information security projects. Shi-Ming and co-authors, for instance, proposes a model based on the balanced scorecard (BSC) framework to help manufacturing companies in Taiwan assess the performance and benefits of their investments in information security projects (Shi-Ming, Chia-Ling et al. 2006). However, traditional accounting performance measures such as Return on Investment (ROI), or economic measures such as Internal Rate of Return (IRR) on investments seem to be not appropriate to determine *ex ante* security investments (Gordon and Loeb 2002). Nonetheless, it is possible to establish the relationship between information security investments and firm market value *ex post*. Chai and co-authors examines the relationship between firm's security investment announcements and a firm's value on the stock market (Chai, Kim et al. 2011). They find that announcements of information security investments bring positive abnormal returns when the firm announces it will be selling or offering new products or services with information security elements. Investors also respond more positively to security investment

announcements after the enactment of the Sarbanes–Oxley Act. In contrast, Campbell and co-authors find highly significant negative market reaction for information security breaches involving unauthorized access to confidential data (Campbell, Gordon et al. 2003). Accordingly, Acquisti and co-authors finds a negative and statistically significant impact of data breaches on a company's market value on the announcement day for the breach (Acquisti, Friedman et al. 2006). The large majority of studies in the area of market reactions to security incidents report statistically significant results of the negative impact of security events on firm stock price (Spanos and Angelis 2016).

As a result, negative market reactions can be one of the reasons which motivate a certain type of organisations to allocate resources to information security projects. Besides producing economic losses, data breaches can generate disputes with customers and trigger costly enforcement actions by regulators. Institutional pressures, such as regulatory changes, and internal security needs assessment seem in fact to significantly explain the variation in organisational investments in information security measures (Cavusoglu, Cavusoglu et al. 2015). However, organisational restructuring events, such as merger and acquisitions, can also contribute to reshape the prevalent security culture and to drastically change organisational priorities (Dhillon, Syed et al. 2016). Other organisational characteristics can also play a role in resource allocation to information security projects. For instance, information security might be significantly down the list of investment priorities for small and medium enterprises which do not process large amounts of data (Goucher 2011).

Corporations differ in the type of dataveillance and privacy strategies they decide to adopt (Clarke 1996). They may deny the strategic significance of security for the company, or wait for some security breach to happen and react to it. On the other side, they might take a more proactive approach and prepare for more likely contingencies affecting data security. Finally, they might try to differentiate themselves from their competitors by implementing tools and practices to safeguard information security and data privacy. If organisations can react differently to the side effects of data use, the same boundaries of what is considered lawful and what is considered

allowable use of personal data are blurred and changing. Therefore, the identification of those elements that motivate companies to protect, or not to protect, data beyond compliance could help all actors engaged in the data protection debate to dialogue and offer more effective, and reassuring responses on the matter. Since information security is a fundamental component of data protection, exploring the reasons behind information security investment decisions may contribute to shed light on the degree of compliance with data protection principles. Therefore, this study will offer an exploratory analysis of the potential reasons behind investing in information security.

3.9.1 Privacy preserving measures

Data breaches can be caused by a number of different technical, procedural or organisational failures. The more sensitive and valuable data are, the more they are at risk and need to be protected from cybercriminals and other malicious agents. Several procedures and technologies exist to ensure data are protected from abuse. The aim of this and following sections is to present common strategies and technologies adopted within organisations to protect data from abuse.

This section focuses on *Privacy Enhancing Technologies* (PETs). PETs are a result of the success of the 'privacy as confidentiality' paradigm predominant in computer science (Gürses and Berendt 2010). PETs represent "a system of ICT measures protecting informational privacy by eliminating or minimizing personal data thereby preventing unnecessary or unwanted processing of personal data, without the loss of the functionality of the information system" (van Blarkom, Borking et al. 2003: p. 33). Anonymity, an objective of PETs, is achieved by decoupling a person's identity from the traces that his/her digital activities leave behind. Once data are anonymized, they are no longer classified as personal data and no longer fall under the protection regime which imposes limitations to data flow. As stated in the data protection directive: "whereas the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable" (EC/46 1995: p. 5).

PETs protect personal data in a number of ways. It can protect privacy (a) by limiting the ability of others to discern the identity of a particular person, for example, by means of an anonymizing browser (i.e. subject-oriented PETs); (b) through the use of a particular technology, like anonymous e-cash (i.e. object-oriented PETs); (c) through the safeguard of transactional data, by means of, for example, automated systems for destroying transactional data (i.e. transaction-oriented PETs); or (d) through the creation of areas of interaction where the identity of the subjects is never recognizable, like in the case of anonymous remailer systems (i.e. system-oriented PETs). An example of PET is the Platform for Privacy Preferences (P3P) underway at the World Wide Web Consortium (www.w3.org/P3P/). In that case, data protection principles conceived to shape bureaucratic routines have been translated into technical practices, by means of software that automatically negotiate privacy licenses.

This type of solution seems not to have experienced commercial success in the end-consumer market, though. Although this evidence might signal the absence of a significant demand for those products, consumers are more willing to buy from more privacy protective merchants, even when that may entail paying modest price *premia*; which means that privacy protection may be revenue enhancing (Tsai, Egelman et al. 2008). Finally, as PETs are adopted on a voluntary base, they cannot ensure the same level of protection to everyone.

PETs are nothing more than the technological manifestation of a set of norms and principles about data confidentiality and secrecy of communications envisioned by the institutional milieu in which those technologies are embedded. As said before, data protection principles can be enacted through technical solutions of legal arrangements. Consensual data are obtained, for example, by means of either *opt-in* or *opt-out* contractual agreements. In the first case individuals must give their explicit consent to the collection and use of their own data. In the second case consent is implicit if it is not explicitly withdrawn. When we agree about the terms and conditions of a privacy policy, by signing it or ticking a box, we grant our consent to the gathering and processing of our personal data. Although individuals should be informed when data about them is gathered, huge amount of *trace* data produced daily are non-consensual (UN 2004).

3.9.2 The role of privacy professionals

The complexity and variety of norms enforced across different jurisdictions, and the restless movement of information across borders, have created a strong demand for privacy professionals to be hired as firm employees or external consultants. With respect to the interpretation of the law, private firms, especially financial service providers, have tried to cope with the fragmented regulatory privacy landscape through the establishment of international legal teams to manage relationships with national data protection agencies, the implementation of international information systems standards, and ongoing privacy training for compliance officers (Frasher 2013).

Privacy professionals seem to play a more proactive role in the US, where they are involved in defining the organisation's information management strategy, and a more reactive and bureaucratic role in the EU, where they primarily ensure compliance with the law (IAPP 2010). Privacy professionals help companies to adapt to legislation once it is in place or can be hired to assess the quality of current privacy policies and to suggest improvements (Connolly 2008, Robinson, Graux et al. 2009). As showed in table 9, most large corporations have dedicated privacy departments.

The Chief Privacy Officer (CPO) function helps organisations include privacy in both top-down activities, such as employee training, as well as in the communications with the board of directors. They sit at their firms' senior management level, and their activities largely involve strategic, rather than purely operational, issues. They spend a great amount of time assessing the state of dynamic privacy norms by interacting with external stakeholders including regulators, advocates, and professional peers. Privacy is also operationalized through a distributed network of employees which includes both dedicated privacy professionals and specially trained employees within business units.

> These distributed mechanisms, on the one hand, extend the reach of the CPO into the firm, creating a bidirectional system that communicates privacy objectives downstream

while facilitating the identification of new issues and escalation upwards. On the other hand, this architecture enhances the legitimacy and effectiveness of the privacy function by both engaging the business units in defining and tailoring privacy's operationalization within specific corporate environments and also placing responsibility for compliance with these agreed upon business-aligned privacy objectives with the senior executives within each unit (Bamberger and Mulligan 2011: p. 479-480).

Table 9. US Fortune-500 companies with CPOs

List of Companies with Chief Privacy Officers (CPOs)		
Aetna Inc.	Electronic Data Systems	Pfizer Inc.
Agilent Technologies Inc.	Eli Lilly and Co	Principal Financial Group
American International Inc.	Express Scripts	Procter & Gamble Co
Ashland	Exxon Mobil Corporation	Qwest Communications
AT&T Corp	Ford Motor Company	Sprint Nextel
Automatic Data Processing	Gateway Inc.	SPX Corporation
AutoNation Inc.	General Electric Corporation	Sun Microsystems Inc.
Bank of America Corporation	General Motors Corporation	Sunoco Inc.
Berkshire Hathaway	Goldman Sachs Group	Supervalu Inc.
Brunswick Corporation	Hewlett-Packard Company	The Charles Schwab
Cardinal Health, Inc.	Home Depot Inc.	Corporation
Caremark Rx, Inc.	Intel Corporation	The Walt Disney Company
Cendant Corporation	International Business	TIAA-CREF
Chevron Texaco Corporation	Machines Corporation	Time Warner Inc.
Cisco Systems, Inc.	J.P. Morgan Chase & Co	US Bancorp
Citigroup Inc.	KeyCorp	Unisys Corporation
Comcast Corporation	McKesson Corporation	UnumProvident
ConocoPhillips	Merck & Co Inc.	USAA
Countrywide Financial Corp	MetLife Inc.	Verizon Communications
Deere & Company	Microsoft Corporation	Wachovia Corporation
Dell, Inc.	Nash Finch Company	Wall-Mart Stores, Inc.
Delta Air Lines Inc.	Nationwide Mutual Insurance	Washington Mutual Inc.
Eastman Kodak Company	NCR Corporation	Whirlpool Corporation
	Oracle Corporation	Wyeth

Source: Author's elaboration of (Shalhoub 2009).

Thus, besides national or sector-specific regulations, over the past years several legal and technological solutions have been proposed and developed to tackle privacy problems. On one side, opt-in/out contractual terms have spread across all sectors to meet customer privacy expectations (Sovern 1999, Milne and Rohm 2000). On the other side, several privacy enhancing technologies have been developed, though without reaching high commercial success in the end-consumer market (Acquisti 2010). A comprehensive assessment of the effectiveness, proliferation and complementarities of these solutions among companies processing personal

data under the European data protection legal framework still needs to be achieved (Gutwirth, Poullet et al. 2011, van der Sloot 2014).

3.10 Limitations of the current European data protection regulatory regime

Legal practitioners and privacy activists have been pointing out the limits and ambiguity of the way the Directive has been transposed by EU member States and its profound lack of harmonisation (Robinson, Graux et al. 2009, Art29 2010, Art29 2010a, Art29 2010b). Originally the 1995 Data Protection Directive was meant to ensure the unrestrained flow of data within the European Union, which was characterized by differing data protection regimes (Shaffer 2000). However excessive member state bargaining produced a Directive drafted as a combination of national laws, rather than as a consistent body of norms and procedures formulated *ex novo*. The target of protecting data-subject rights without creating any disruption to business operations, proclaimed in the Data Protection Directive 1995, seems far from being achieved. There are difficulties in determining *whether* EU data protection law applies to processing of personal data (e.g. cases of exemptions and exceptions), and, if this is the case, what law would be applicable, how this law should be applied and how its applications could be enforced (Korff 2010). The multidimensional legislative nature of the problem, constantly challenged by rapid change in the techniques and procedure to gather and process personal data, contributed to the expansion of a complex ramification of norms, in the form of amendments or new legislative layers.

Although the evolution of the ICT sector and the increasing global integration of markets seem to leave law makers always a step behind, in 2009 the European Commission decided to assess the state of compliance with the Data Protection Directive 1995, its limits and room for improvement. The idea was to overcome the limits identified in the current regulatory framework and increase integration of data protection practices across jurisdictions. The attempt has been considered crucial for both the development of contractual agreements between business partners operating globally and the expansion of those ICT infrastructures that form the modern digital economy (Shires 2011, EC 2014). The revision process started with a round of consultations with various

stakeholders (Robinson, Graux et al. 2009), from industry representatives (ISFE 2010), privacy advocates (PI 2009), to third countries representatives, like the US data protection ambassadors in Brussels. These actors came to the conclusion that the Directive maintained a certain degree of ambiguity in the terminology adopted, room for power in the large set of exceptions mentioned, as well as in the guidelines for implementation, which, largely assigned under the competence of European member states, had created over time an heterogeneous and unpredictable interpretation of the law (Simitis 1994).

As a result, on January 2012 the European Commission prosed a comprehensive reform of the EU's 1995 data protection rules meant to harmonise data protection rules across EU member states (DG-Justice 2012). The legal instrument chosen was a 'regulation' instead of a 'directive', directly applicable to all EU member states without a need for national implementing legislation. In March 2014, the European Parliament adopted a revised version of the draft regulation – after 3999 amendments generated by intense lobbying. By the summer of 2014 general agreement on various chapters (I, IV, V and IX, which deal with specific rules for public authorities and other special sectors, obligations for the data controller and processor, and international data transfers) was achieved. In the spring of 2015 both the European Parliament and the Council were aiming at opening 'trilogue' negotiations on the final version of the regulation by the summer, as well as finalising the overall legislative process by the end of 2015. If finally approved, the Regulation will be enacted, after a two years transition period, in every EU Member State (LIBE 2015).

Since the environment in which the traditional privacy principles are now implemented has undergone significant changes (OECD 2013), several unspoken questions remain on how to apply data protection principles in big data environments. For instance, "using personal information in big data analytics may not be permitted under the terms of the original consent as it may constitute a secondary use—unless consent to the secondary use is obtained from the individual" (Cavoukian, Stewart et al. 2014: p. 11). Although new technological developments and global economic dynamics may pose some difficulties to the application of these principles, the core data protection principles enclosed in the EU Data Protection Directive have been included in the

proposed General Data Protection Regulation as they are still widely accepted and considered by the majority to be still valid today (DHS 2008, Art29 2009, ECDP 2010).

3.11 Conclusions

As presented in Chapter Two and Chapter Three, the creation, analysis and protection of personal data are strictly intertwined phenomena. As law makers are equally interested in protecting citizens' rights, as well as in fostering economic growth and innovation, an in-depth analysis of the way data protection principles are implemented within organisations will help us better understand the drivers behind the adoption of good information security and privacy practices and procedures. The emphasis on the idea of privacy, part of the organisation's ethical culture, is also expected to help us understand the complex path toward the transformation of abstract privacy principles into actual information security procedures.

The central role played by the privacy regulatory regime has also been acknowledged within this chapter. Enforcement actions such as sanctions or prosecutions, and consequent litigation risks, are instruments adopted by regulators to force organisations adopt minimum information security measures. Regulation, competition, market dynamics and technological change represent some of the basic leverages shaping the overall organisational strategy, and the organisation's information security management strategy in particular. Data protection laws can influence organisational decisions in many ways and at different levels. To better understand the relationship between big data and data protection the next chapter will draw insights from previous academic studies in order to investigate the impact of the privacy regulatory regime, and of the level of analytical sophistication an organisation has achieved, on the degree of compliance with data protection principles.

CHAPTER FOUR

Research framework and hypotheses

4.1 Introduction

The function of this chapter is threefold. First of all it summarises the themes identified in the previous two chapters, which refer to organisational information management practices with respect to both big data analytics and data protection. Second, it highlights the knowledge gap within the privacy studies literature and presents the research question. Finally, the chapter exhibits a set of propositions concerning how the concepts and ideas identified address the research question.

4.2 Research gap

Information privacy is defined as one's ability to control information about oneself (Pavlou 2011). Data protection laws and procedures are meant to safeguard people's data privacy and help data subjects control information about themselves. Scholars have put a big effort into reconceptualising and measuring individuals' privacy concerns and preferences (Clifton, Kantarcioglu et al. 2002, Solove 2004, Ashworth and Free 2006, Solove 2006, Warren, Bayley et al. 2008). While the study of privacy at individual level has received wide attention, there is a need to expand the investigation of organisational information privacy practices beyond the analysis of privacy policies of websites (Ryker, Lafleur et al. 2002, Jensen and Potts 2004, Sheehan 2005) and understand the problems companies may face in their attempt to comply with fair information practices (Bélanger and Crossler 2011). Privacy studies also seem to lack a theory to explain firms' information privacy behaviours (Greenaway and Chan 2005). Studies which examine information privacy policies posted to firms' websites across industry sectors and jurisdictions (Ryker, Lafleur et al. 2002) provide limited insight into the complexity of the

information privacy phenomenon within organisations (Milne and Culnan 2002). Thus, there is a need to better understand the kind of procedures organisations implement to manage users' privacy concerns.

Exploring the way organisations treat personal information is of paramount importance as it influences people's data sharing behaviour. The fact that data are processed by employers, insurance companies, law enforcement agencies, or the Internal Revenue Service generates different level of concerns across people (Stone, Gardner et al. 1983). Moreover, the perceived fairness of corporate information practices can decrease consumer privacy concerns (Culnan and Bies 2003). As a result, there is a need to understand organisational data privacy practices and how they match individual privacy concerns (Bélanger and Crossler 2011).

This research contributes to previous studies, which have investigated the effect of the regulatory approach to information privacy on the corporate privacy management environment (Culnan 2000, Milberg, Smith et al. 2000), by paying attention to the effect that becoming analytically sophisticated (Davenport and Harris 2007, Davenport 2014) has on organisational data privacy decisions. Analytically sophisticated organisations seem to invest in, and take advantage of, big data analytics and also understand the importance of protecting people's information privacy (Davenport and Dyché 2013); thus, analytical sophistication and data protection might complement, rather than clash, with each other (Schermann, Hemsen et al. 2014). As the way in which employees understand and enact the law plays a fundamental role in determining legal compliance (Ball 2010, Dibb, Ball et al. 2014), effects of the privacy regulatory regime on the creation of an internal privacy culture will also be explored. Finally, the surveillance studies literature may help uncover the unintended consequences and negative dimensions related to the growing concentration of digital data (Lyon 2001, O´Hara and Shadbolt 2008, Zureik, Harling Stalker et al. 2010).

In particular, the concept of dataveillance (Clarke 1988, Degli Esposti 2014, van Dijck 2014) can help us understand the relationship between the way information privacy is currently conceived and the proliferation of targeted analytics and customer profiling.

Surveillance scholars tend to emphasise discriminatory and worrying aspects of data usage (Stanley 2004, Amoore and DeGoede 2005, Ayres 2007, Payne and Trumbach 2009, Gandy 2012), even though they acknowledge that digital surveillance bring both risks and benefits (Lyon 2001). Data protection and privacy laws are meant to limit dataveillance performed by governments and corporations (Bankston and Soltani 2014). Legislative attempts to safeguard information privacy are criticised by both surveillance scholars (Gilliom 2011) and the proponents of the corporate perspective on consumer privacy, who argue that any restrictions placed on the private firms' ability to access personal information about consumers compromises their ability to operate efficiently in the marketplace, and thus impedes its ability to fulfil its social responsibility of creating economic growth and development for society (Lester 2001, Culnan and Bies 2003).

Although considerations related to data protection and technology trends like big data have important policy implications, there is a lack of empirical studies on the matter (Tene 2012). The study of information privacy at organisational level has received limited attention, despite being considered an organisational ethical imperative (Mason 1986, Smith 1993, Culnan and Smith 1995); no study has also tested claims based on previous qualitative studies which suggest that analytically sophisticated organisations understand the importance of respecting data privacy and are capable of transforming it into an organisational value (Davenport, Harris et al. 2010).

Finally, there is a need in privacy studies to investigate the kind of practices and procedures organisations adopt to protect information privacy outside the context of the United States (Bélanger and Crossler 2011). In this respect, Europe represents a very interesting context and an opportunity to explore the effect of the privacy regulatory environment on corporate decisions. The EC Data Protection Directive 1995 stresses both the importance of safeguarding privacy as fundamental human right, as well as the importance of safeguarding data transfer in the digital

era. Although the legislation recognises, in principle, the privacy-transparency trade-off, in practice, little is known about how data-controllers, usually private companies, interpret and implement data protection policies while they are also trying to achieve their business objectives (Mantelero 2014, Ciriani 2015). Past research has been focused on understanding privacy from a broader legal or societal perspective (Shires 2011, Blume 2012), without paying attention to how the speed of technological change in database management affects the actual implementation of the law and the stipulation of optimal contracts between business partners operating in different jurisdictions (Brown 2010, Christensen and Etro 2013).

This research hopes to address some of these issues by focusing on the research questions presented in the next section.

4.3 Research questions

Regulatory changes meant to safeguard individual information privacy influence business decisions and overall organisational information management strategies. Because of changes in the data protection regulatory landscape and in digital market dynamics the need to pay more attention to the study of privacy at organisational level has become more urgent. According with the gaps identified in the previous section, this study will attempt to answer the following research questions.

Research Question One: How does the data protection regulatory regime influence enterprise data protection and data management decisions?

Research Question Two: How does the level of analytical sophistication an organisation has achieved influence enterprise data protection and data management decisions?

Since this study focuses on the interplay between big data and data protection the research questions refer broadly to ‘organisational information management decisions’. The rest of this chapter is devoted to review the academic literature in search of insights to answer these questions. The next chapter presents the definitions of all constructs used in the study and their

empirical operationalisation. At the end of this chapter the theoretical model, which shows the articulation and directionality of relationships among constructs, will be presented.

4.4 The phenomenon under study

This study wants to shed light on the way data protection and big data interact with each other. Accordingly, the first research question focuses on effects of the data protection regulatory environment on data handling decisions, while the second question pays attention to technological aspects, such as the extent to which the organisation relies on big data analytics to pursue its objectives.

In order to answer both questions, the study proposes to examine the *corporate privacy management environment* (Milberg, Smith et al. 2000), by investigating a new multi-dimensional construct: *the Degree of Compliance with Data Protection Principles, particularly Compliance with Data Controllers' Obligations (DPP)* and *Respect of Data Subjects' Rights (DSR)*, as defined in the EU Data Protection Directive 1995 and corresponding national legislations. The proposed construct will help us assess the extent to which companies implement widely recognised data protection principles, as part of their internal policies, practices and procedures. The two constructs are multidimensional and include the following elements. The construct **Compliance with Data Controllers' Obligations (DPP)** can be defined as follows:

> The organisation keeps data complete, accurate and up-to-date, tries to collect the minimum amount of data necessary to fulfil a specific objective and shares individuals' data only with authorised third parties. Data are also deleted once the objective for which they have been collected is achieved. Within the organisation, strong security measures protect data from unauthorised use. There are also procedures in place to compensate individuals in case data were lost, manipulated or stolen, and to sanction those who use or handle personal data inappropriately.

The construct *Data Controllers' Obligations* is formed of seven core dimensions.

1) *Data quality* refers to the extent to which personal data are kept in a complete, accurate and up-to-date form.

2) *Purpose specification* refers to the practice of collecting the minimum amount of data necessary to fulfil a specific objective.

3) *Retention* refers to the practice of erasing data once the objective for which they have been collected is achieved.

4) *Data transfer* refers to procedures for sharing individuals' data only with authorised third parties.

5) *Data security* refers to the security measures adopted by organisations to protect data from unauthorised use.

6) *Accountability* refers to the presence of sanctions for those who use or handle personal data inappropriately

7) *Redress* refers to the existence of procedures to compensate individuals in case data were lost, manipulated or stolen.

The construct **Respect of Data Subjects' Rights (DSR)** can be defined as follows:

> The extent to which an organisation manages to ensure that individuals are fully informed about all aspects related to the processing of their data; it ensures that individuals are also asked to give their explicit consent to personal data processing; and that there are procedures in place to let individuals rectify inaccurate data and satisfy their requests to end the processing of their personal data.

Thus, data subjects enjoy four fundamental rights related to the right to data protection; these rights are: the right of notice; the right to consent; the right of access; and the right to object (see section 3.3). Accordingly, the construct *Respect of Data Subjects' Rights* is composed of four dimensions.

1) The *right of notice* refers to the fact that individuals must be informed about all aspects related to the processing of their data.

2) The *right to consent* refers to the fact that data controllers must obtain explicit consent from individuals before processing their data.

3) The *right of access* refers to the existence of procedures to let the individuals rectify inaccurate data.

4) The *right to object* refers to data controllers' ability to satisfy individuals' requests to end the processing of their data.

Different factors may play a role in shaping an *organisational privacy behaviour*, defined as the way "firms treat their customers' personally identifiable information" (Greenaway and Chan 2005: p. 172). In the following sections the researcher will try to identify these factors and highlight how they interact with each other.

4.5 Answering the first research question

Little is known about the way privacy regulatory regimes influence organisations across jurisdictions and how what it constitutes lawful use of data might influence the development of firms' business models. Assessing the effects of privacy laws on business is especially problematic because of the speed of technological innovation, which certainly contributes to limit the effectiveness of any technology-specific legal instrument. For instance, the use of Web bugs, also known as clear GIFs, which offers a way to track not only Web page navigation but also the opening of individual e-mail message, may make cookie legislation obsolete right after its publication (Miyazaki 2008). In Europe owners of databases enjoy the right to object to the copying of substantial parts of their database, even if data is extracted and reconstructed piecemeal (EC 1996). Given that only a limited number of cases have been ruled by the European Court of Justice over the past ten years under this regulation, several commentators take this

regulation as another example of the large ineffectiveness of legal instruments in the attempt to regulate information management issues.

To complicate the issue further, privacy laws vary considerably across jurisdictions as they reflect historical, political, and cultural differences. It has been empirically demonstrated that regulatory approaches to information privacy differ according to cultural values and privacy concerns, and that the demand for stricter regulatory measures increases when private organisations are perceived to treat personal data improperly (Milberg, Burke et al. 1995, Milberg, Smith et al. 2000). Countries where people exhibit higher levels of anxiety, stress, and concern for information privacy and security, not surprisingly are also those where people express preferences for clear written rules and regulations (Hofstede 1980, Hofstede 1991, Milberg, Burke et al. 1995).

With regard to Europe, commentators claim that the 1995 Data Protection Directive has not functioned as an entry-barrier for foreign companies (Kane and Ricks 1988, Connolly 2008) as it was initially foreseen (Samiee 1999), and the US government has also not improved its privacy-enhancing procedures in order to protect trading flows as originally expected (Shaffer 1999, Shaffer 2000, Brookman 2015). In addition, the EU Directive has not been effective in tackling the problem of dealing with multiple jurisdictions, as proved by the Safe-Harbour US-EU agreement (Kobrin 2004). Other unilateral or multilateral agreements adopted as a result of the introduction of the 1995 Data Protection Directive have only saved appearance without changing actual practices (Connolly 2008). It is also difficult to assess the way the Directive has influenced operations and costs of European subsidiaries of multinational corporations (Samiee 1984).

In the context of European data protection law two specific aspects must be taken into consideration, which are the impact of law on organisations' data protection practices and the impact of the law on firms' ability to innovate and generate economic value. The effects of regulation over firm data management practices and firm operations cannot be assessed *a priori*, but change depending on how laws are formulated and to the extent to which they influence

business objectives and operational functions. Regulation can also be only partially implemented, misinterpreted, or opposed. For this reason, we will take into consideration two particular properties of data protection legislation: its clarity and its enforceability.

We can take as an example the enduring resistance to the US *Uniform Computer Information Transactions Act* (UCITA). The UCITA case is an example of a type of data protection legislation which has been judged by the majority unfeasible and counterproductive and has never become effective because of the limits of adopting a property-right approach to the theme of data protection (Samuelson 2000). This case shows how contradictory legislative and self-regulatory initiatives may generate a growing climate of uncertainty wherein consumers and firms are forced to operate in sub-optimal conditions (Acquisti 2010). As the same author explains:

> This uncertainty is costly in itself, in that it forces data subjects and data holders to invest resources into learning about the admissibility of a given data practice. It also creates costly second order effects, in that it may lead both data subjects and data holders to inefficiently under- or over-invest in data protection. Similar costs arise for Internet companies that operate worldwide and need to conform their services to differing local standards of privacy protection. (Acquisti 2010: p. 14).

The reform of the Data Protection Directive 1995, which started with the publication of the first draft of the proposed General Data Protection Regulation on the 25th of January 2012, has opened a long period of uncertainty on the future of European data protection law. It has also triggered ongoing discussions between supporters and detractors of the proposed new regime (Blume 2012). Commentators in favour of the new regulation believe that data protection law had to be strengthened in order to safeguard data subjects' rights (Reding 2012, CL&SR 2013). Another claimed advantage is the harmonisation in the field of data protection which is expected to avoid the business costs associated with dealing with 28 different national privacy legislations in Europe (Dix, Thüsing et al. 2013). Because of the emphasis given to new concepts such as data-protection-by-default and security-by-default, the proposed Regulation is also expected to strengthen the

European ICT industry and to increase the demand for Europe-based cloud services (Christensen and Etro 2013). Nonetheless, the Regulation has also been criticised for its inaccessible language and its limited capacity to empower data subjects (Blume 2014).

Organisations tend to adopt four strategies to cope with regulatory uncertainty; these are: avoidance; reduction; adaptation and disregard. Reduction seems to be the most common strategy (Engau and Hoffmann 2011). In the case of response to post-Kyoto regulatory uncertainty, 75% of respondents (n = 112) said that they tried to reduce regulatory uncertainty by systematically searching for additional information, by focusing on specific issues in their business environment and by engaging in current policy making processes to simplify the decision making process (Engau and Hoffmann 2011). In addition, firms would be less proactive toward the use of personal data if they perceive that the data protection regulatory regime is opaque and uncertain (Acquisti 2010). The presence of exceptions, legal vacuum, or general disregard of the law might result in a repeated unlawful use of data by firms. Certainly organisations which operate under clear and consistent rules have more chances to respond to normative pressure in a way that increase both their legitimacy and market potential.

Thus, privacy laws may influence business decisions not only when it poses restrictions on the use of data, but also when it does not state clearly its terminology, objectives and scope. Two complementary aspects must be taken into consideration when assessing regulatory effectiveness: the clarity and predictability of interpretation of the rule of law, and the kind of sanctions and enforcement powers of those agencies overseeing the application of the law. Accordingly, the construct **Data Protection Regulatory Regime (REG)** is defined as follows:

> the extent to which data protection law is enforced in a consistent, reliable and predictable manner and data protection authorities have the power and the resources to impose serious sanctions if data are processed unlawfully.

The next two subsections will explore how data protection regimes characterised by clarity and strong enforcement powers may influence organisational decision-making in areas such as data privacy and data usage.

4.5.1 Effects of regulation on the organisational privacy culture

Laws and regulations need to be interpreted and enforced in a consistent and reliable manner in order to generate a positive, privacy-friendly institutional environment for private firms and public entities. In the case of data protection law, the objective is to ensure that organisations, which process personal data, respect data subjects' rights by following some basic data protection principles, which have been presented in Chapter Three. The constructs *Respect of Data Subjects Rights* (DSR) and *Compliance with Data Controllers' Obligations* (DPP), discussed in section 4.4, summarise these ideas.

Organisations can enhance their privacy programs by creating a culture of integrity that combines a concern for the law with an emphasis on managerial responsibility for an organisation's privacy behaviour (Culnan and Williams 2009). Thus, the construct *Organisational Privacy Culture* (PRV) is meant to capture the extent to which privacy is seen to be part of a company's culture (Greenaway and Chan 2013). Organisational, or corporate, culture is expressed in the collective values, norms and knowledge within an organisation, which affect the behaviour of employees (Schein 1985). Accordingly, an **Organisational Privacy Culture (PRV)** can be defined as:

> the extent to which privacy represents a distinctive brand feature and a core value, central to the organisational culture, which implies that remarkable human and financial resources are devoted to secure information.

European and national data protection laws are meant to foster the organisational privacy culture through the implementation of different kinds of procedural and technical measures. As an example, previous studies have shown how the Italian Data Protection Authority contributed to the development of requirements engineering methodologies able to capture security obligations

by imposing on public administration the requirement of implementing minimal precautionary security measures, such as the adoption of authentication and authorisation systems (Massacci, Prest et al. 2005).

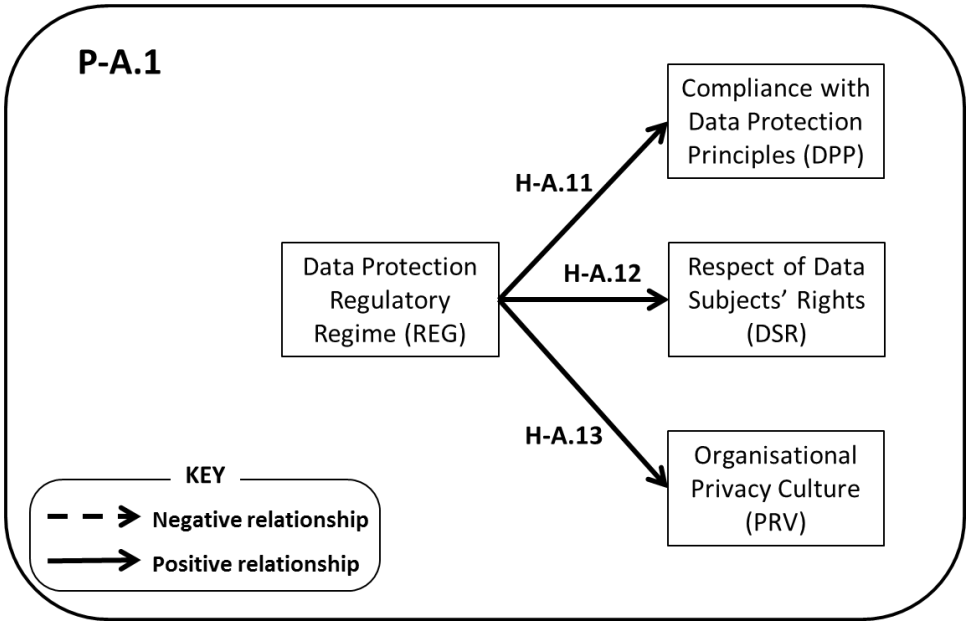
The fact that data protection agencies have the power to impose serious sanctions can also influence organisational data handling procedures. A study undertaken by the UK Information Commissioner's Office (ICO), based on in-depth telephone interviews with 14 organisations which had received Civil Monetary Penalties (CMPs), and 85 'peer' organisations from similar sectors who had not received any sanction, shows that organisations tend to review or change their data protection practices and policies as a result of hearing about CMPs being issued to other organisations. In addition, organisations, which have been issued with a fine, take their data protection obligations more seriously, with revised practices and policies, and increased staff training (ICO 2014).

In addition, because they have more to lose, companies which make revenues out of data collection and processing are expected to implement data protection practices beyond compliance (Culnan and Williams 2009). In fact, the implementation of more conservative data protection measures would probably reduce the risk of data breaches, which would imply lowering the economic and reputational costs arising from a loss of data (Acquisti, Friedman et al. 2006). Thus, both the fact that law enforcement agencies have the power to impose considerable sanctions and the fact that courts interpret the law in a consistent and predictable way can contribute to shape organisational decisions in the area of data protection. These insights lead the researcher to formulate the following proposition:

> ***Proposition A.1:*** *The less permissive and more reliable the data protection regulatory regime (REG), the more likely it will be that organisations develop an internal privacy culture (PRV), respect data subjects' rights (DSR) and comply with data protection principles (DPP).*
>
> ***Hypotheses***
>
> - *H-A.11 - REG will be positively associated with DPP.*

- *H-A.12 - REG will be positively associated with DSR.*
- *H-A.13 - REG will be positively associated with DSR.*

Figure 5. Proposition A.1 and corresponding hypotheses



4.5.2 Effects of regulation on data analysis procedures

The modern microeconomic theory of privacy shows that the protection of personal privacy through regulatory measures can increase aggregate welfare as much as the interruption of data flows can decrease it (Acquisti 2010, Brandimarte, Acquisti et al. 2012). Lack of consumer data and fear of possible legal reprisals following the collection or processing of consumers' data may hamper service and product innovation. In general, there is agreement about the positive effects of data availability on firm results across business functions (e.g. sales, advertising, inventory and marketing; R&D and product/service development; risk and information management). For instance, the availability of consumer data contributes to the development of more sophisticated marketing solutions, which might imply higher market penetration, marketing returns on investments and profits (Thomas and Maurer 1997). In contrast, a reduction in the information flow, due to regulation or other causes, may affect firm internal efficiency and innovation capabilities. Differences in legal restrictions applied to the use and collection of personal data

might enable, or hamper, alternative technological paths, which in turn may cause changes in the proliferation of firms whose business model is highly dependent on data.

Although data protection laws can improve organisations' information security and privacy procedures, their provisions can create obstacles to the use of data and to the creation of new services. For example, it has been predicted that the effects of mandatory opt-in privacy policies on companies like multinational financial institutions would be disastrous (Staten and Cate 2003). The authors of this study claim that, while both opt-in and opt-out give consumers the final say about whether their personal information is used, an opt-out system sets the default rule governing use of personal information to 'free flow,' while an opt-in system sets the default rule to 'no information flow.' Thus, switching from an opt-out to an opt-in system would "raise account acquisition costs and lower profits, reduce the supply of credit and raise credit card prices, generate more offers to uninterested or unqualified consumers and raise the number of missed opportunities for qualified consumers, and impair efforts to prevent fraud and identity theft" (Staten and Cate 2003: p. 783).

Security and privacy concerns related to big data and an IT infrastructure that is accessed through remote locations, for example, via a data cloud and services hosted in the cloud, have presented a significant barrier to the adoption of big-data approaches (Roski, Bo-Linn et al. 2014).

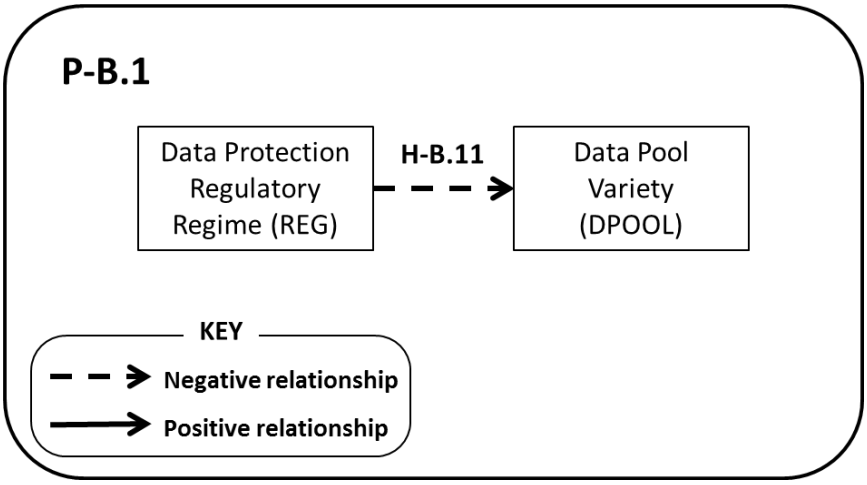
The European Data Protection Directive 1995 contains provisions which could have prevented the creation of big data repositories. These provisions set strict requirements for companies which want to transfer data between countries (Schwartz 1994). The purpose limitation principle, part of the same Directive, poses also restrictions to data accumulation and retention. According to this principle, data can be processed only for a specified, explicit and legitimate purpose; and any further processing must be compatible with the original purpose for which the personal data were collected. In April 2013, the Article 29 Working Party adopted Opinion 203 (Articles29 2013), which elaborates on the purpose limitation principle set out in Article 6(1)(b) of the EU Data Protection Directive 95/46/EC. The Opinion advises organisations to adopt opt-in consent

procedures in the case big data technology is used. The compatible reuse of data has always to be assessed on a case-by-case basis.

These insights lead the researcher to formulate the following proposition:

> ***Proposition B.1****: The less permissive and more reliable the data protection regulatory regime (REG), the less likely it will be that organisations indiscriminately collect and analyse a large variety of data (DPOOL).*
>
> ***Hypothesis:***
>
> - REG will be negatively associated with DPOOL.

Figure 6. Proposition B.1 and corresponding hypothesis



Economic theory considers that privacy regulation can be beneficial, or deleterious, for organisations depending on several factors, such as its content, context of implementation, and privacy preferences of data-subjects. Clear legal requirements can shape the way organisations use data to generate revenues or to structure business process in such a way to comply with regulation and avoid costly enforcement actions by regulators. The 'consent-based' approach, commonly foreseen by privacy laws, may alter the competitive structure of data-intensive industries by imposing transactions costs which oppress small firms while benefitting firms which have already gained market predominance (Campbell, Goldfarb et al. 2015). Privacy law can benefit organisations which have already implemented privacy-preserving procedures or can

create entry barriers which advantage incumbents (Dean and Brown 1995). In general firms tend to adapt to legislation and to find lawful ways to apply technology. In the specific case of big data analytics, the technology can be compatible with the enactment of basic data protection principles.

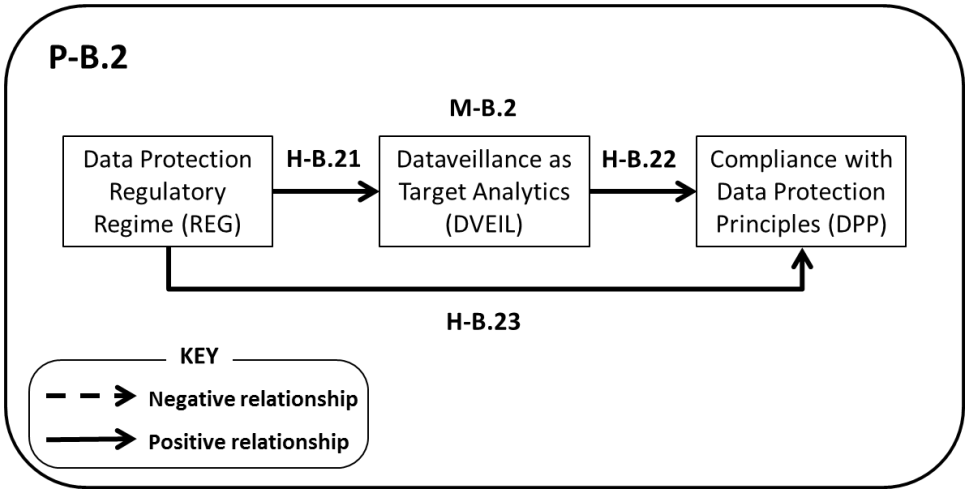
> We believe that it is entirely possible to achieve privacy in the Big data era. We can protect the privacy of personal information while using data analytics to unlock new insights and innovation to move our organisations forward (Cavoukian, Stewart et al. 2014: p. 2).

De-identification, for instance, allows organisations to comply with data minimization principles (Cavoukian and El Emam 2014). Using proper de-identification techniques and re-identification risk management procedures remains one of the strongest and most important tools in protecting privacy. Advances in Biometric Encryption and the application of a Privacy-by-Design approach can also lead to deployment of privacy-protective and secure biometric systems (Cavoukian and Stoianov 2014). It is also possible to envision flexible ad-delivery frameworks which contemporarily maximise ad-relevance, privacy, and efficiency in a single system where personalisation is done jointly by the server and the mobile phone (Hardt and Nath 2012).

These insights lead the researcher to formulate the following proposition:

> ***Proposition B.2****: The less permissive and more reliable the data protection regulatory regime (REG), the more likely it will be that organisations use targeted analytics (DVEIL) in a way compliant with data protection principles (DPP).*
>
> ***Hypotheses:***
>
> - *H-B.21 - REG will be positively associated with DVEIL.*
> - *H-B.22 - DVEIL will be positively associated with DPP.*
>
> ***Mediation effect:***
>
> - *M-B.2 DVEIL will mediate the relationship between REG and DPP.*

Figure 7. Proposition B.2 and corresponding hypotheses



4.5.3 Effects of the organisational privacy culture

Stakeholders' increased awareness of the privacy and security risks of data mismanagement strongly contributes to change organisational practices. Game-theoretic approaches show that as consumers become more concerned about their privacy, it is more likely that all firms adopt privacy protection (Lee, Ahn et al. 2011). Privacy protection, when interpreted as fair information practices, can work as a competition-mitigating mechanism when a consumer is a target of competing firms which are trying to personalise their offers (Lee, Ahn et al. 2011). Similar evidences come also from legal studies. A report developed at the Berkeley School of Law shows, through qualitative findings, that breach-notification laws have significantly contributed to heightened awareness of the importance of information security throughout all levels of business organisations and that they have also improved cooperation among different departments within each organisation (Hoofnagle 2007). Thus, the organisational information security and privacy cultures play a very relevant role in ensuring that employees know and respect values and norms which reflect both regulatory requirements and informational characteristics.

It is also important to align security and privacy policies with system requirements (Antón, Earp et al. 2003). Organisations need to ensure that their employees are aware of information security and privacy policy requirements which are encapsulated into regulatory requirements: "[a] culture must be established in which information is protected from risk and the privacy of the

information is maintained" (Da Veiga and Martins 2015: p. 249-250). Employees are highly heterogeneous and the same training and awareness programmes may produce different interpretations and reactions, and differences in the way data protection principles are enacted. For example, a study based in Taiwan shows that female computer professionals are more effective with respect to their male counterparts in regulating their behaviour when protecting other people's personal information privacy (Kuo, Lin et al. 2007).

An information security culture consists of "the manner in which employees perceive and interact with the controls that are implemented to protect information" (Da Veiga and Martins 2015: p. 165). In addition, the information security culture relates to the norms dictating how to handle data in accordance with its sensitivity and visible artefacts, such as encrypted confidential e-mails, shredders for the destruction of confidential documents, annual online information security training, and statistics of the number of incidents related to employee error or negligence (Da Veiga and Eloff 2010).

In terms of the kind of internal guidelines which should be adopted, the examination of sixty-three private sector and ten university technology codes of conduct revealed that ethical values statements, which describe broad moral notions, are more effective than rules-based codes, which identify specific conducts as acceptable or unacceptable (McGill and Baetz 2011). In other words, universal value-based statements are more likely to help embedded ethical principles in the actor's decision making process than purely prescriptive codes of conducts. An ethical code also positively influences employees behaviour (Stevens 2008) and contributes to lower the risk of unethical computer use (Pierce and Henry 1996).

Training and awareness programmes also contribute substantially to build a strong information protection culture. A case study of an international financial institution over an eight-year period across twelve countries shows that training and awareness have a significant positive impact on the information security culture of an organisation (Da Veiga and Martins 2015). The adoption of codes which reflect social expectations for responsible information use, such as fair information

practices (FIPs) and other data protection principles, can help organisations define guidelines for individual rights and organisational responsibilities to address privacy harms (Culnan and Williams 2009).

These insights lead the researcher to formulate the following proposition:

Proposition A.2*: The more organisations foster their internal privacy cultures (PRV), the more likely it will be that they respect data subjects' rights (DSR) and that comply with data protection principles (DPP).*

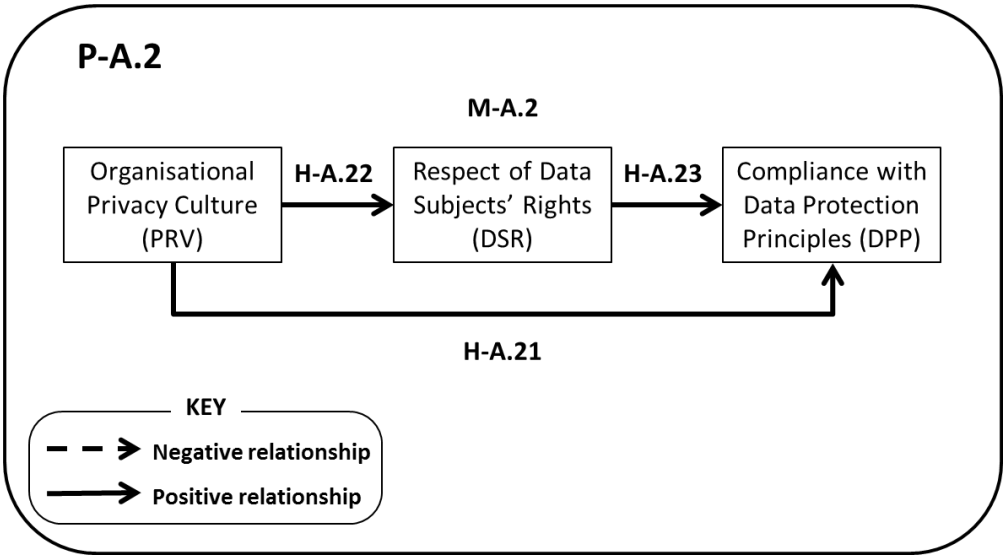
Hypotheses:

- *H-A.21 - PRV will be positively associated with DPP.*
- *H-A.22 - PRV will be positively associated with DSR.*
- *H-A.23 - DSR will be positively associated with DPP.*

Mediation effect:

- *M-A.2 - DSR will mediate the relationship between PRV and DPP.*

Figure 8. Proposition A.2 and corresponding hypotheses



4.6 Answering the second research question

The possibility of analysing the information contained in very large databases, where data are continuously updated and stored in different locations in various formats, is opening a new set of opportunities for businesses to gain efficiency, improve product or service quality and increase sales (Davenport 2014). Most organisations are overwhelmed with data and have to develop the capability to analyse data to make informed decisions and generate business value (Davenport, Harris et al. 2001). Organisations can gain a competitive edge by applying data mining to analyse the information contained in their enormous databases (Harris and Davenport 2007). Enterprises which outperform competitors by using analytics are called analytical competitors (Davenport and Harris 2007, Davenport, Harris et al. 2010). This type of organisations enjoys high levels of analytical sophistication. The term **Analytical Sophistication** (SOPH) indicates that:

> an organisation has a flexible, centralized IT infrastructure to work with data. Data are also accurate, stored in compatible formats and easily accessible and digital data represents a core asset, key to the organisation's business model. Data analytics also represents a distinctive, competitive capability of the organisation: the organisation employs analysts able to mine data and get useful insights and all employees are encouraged to rely on data analytics.

High-performing businesses operating in data-intensive industries, such as the financial, insurance, telecommunication or retail sectors, tend to have a much more developed analytical orientation than other types of organisations (Davenport and Harris 2010). Global knowledge-sharing systems can help firms codify their 'tacit' knowledge and make it available throughout the organisation (Voelpel, Dous et al. 2005). However these systems require important investments in information technologies and the employment of a qualified workforce. Organisations need to have implemented an integrated information management system to gather and make data available to analysts in order to take advantage of big data analytics (Davenport, Harris et al. 2010). Only companies of a certain size seem to make a concerted effort

to maintain and update data necessary for efficacious use of analytics, and place this high on their priorities (Xavier, Srinivasan et al. 2011).

In addition, people with the necessary statistical and mathematical skills to analyse data must also be part of the organisation. Since data scientists are usually people with advanced degree in physics, statistics or computer science, it is difficult to recruit enough people with adequate skills (Davenport and Patil 2012). Yet as the number of companies using analytics increases, it is becoming harder for some companies to gain an edge (Kiron, Prentice et al. 2014). The shortage of analysts is driving organisations to consider outsourcing their analytics activities (Fogarty and Bell 2014). A lack of IT-savvy business people leads to underperforming IT investments (Haggerty 2012). They also often fail to disseminate key insights to employees (Kiron, Ferguson et al. 2013).

> I think the biggest change you see is that everybody in the organisation — whether they are a technical person, a researcher or an engineer, whether they're a product manager, a businessperson, a usual contributor or a manager — everybody has to be data driven (Ferguson 2013: p. 2).

Data availability allows companies to profile customers on the basis of their preferences and according to their actual behaviour. Profiling procedures have furthered companies through more precise targeting, increased effectiveness of advertisement and promotions, and better consumer retention (Lewington, De Chernatony et al. 1996, Loveman 2003, Spangler, Gal-Or et al. 2003, McKechnie 2006, Breur 2007, Jehn-Yih and Pi-Heng 2008, Ogwueleka 2009, Paas 2009, Adams 2010). Employee data can also be used to determine to which employee to assign a certain task or to create models that calculate the optimal number of staff members to deal with customers at the front desk and other service points (Davenport, Harris et al. 2010, Bassi 2011).

Organisations collect large amount of transactional data through different types of IT systems; some examples are: enterprise resource planning (ERP) systems; customer relationship management systems; point-of-sale scanner data in retail stores; web and e-commerce transaction data. These systems help organisations overcome the challenge of information

integration from the point of origin of a product to the points of consumption (Thomas and Jeffrey 2004). The availability of large amounts of different types of data represents a precondition for applying analytics and transforming data into knowledge. For this reason it is important to take into account the nature of the data an organisation is processing. The concept **Data Pool Variety** (DPOOL) will then be used to indicate that:

> A diverse array of data is processed by the organisation. Data such as: geographical location; unstructured data like voice, text or images; people's online behaviours, economic transactions, or individual attributes and attitudes.

Automated decision applications are being used effectively to generate useful solutions in a number of different business areas (Davenport and Harris 2005), such as: *yield optimisation*, which set pricing based on seat availability and the hour or day of purchase; *routing*, which refers to the use of automated filters for sorting cases or transactions by establishing 'priority lanes' to handle orders; *screening*, which indicates automated screening used to perform background checking, identify fraud or authorise payments; *dynamic forecasting*, which perform estimation of customer demand to improve alignment with manufacturing and sales plans; *operational control*, which refers to the analysis of environmental sensors to detect abnormal events and trigger alarms. Thus, analytics can be used by different departments or units within an organisation to pursue several types of goals. To take into account the variety of applications of big data analytics, the concept **Use of Analytics Across Business Functions** (FUNC) can be used to control whether analytics is used to foster marketing, improve security, gain efficiency, to better manage human resources, reduce financial risks, and/or to take better informed strategic decisions. As a matter of fact, big data analytics applications can be equally found in Chinese real estate development and marketing (Du, Li et al. 2014), in the use of social media analytics to foster evidence-based policymaking (Grubmüller, Götsch et al. 2013), or in in the development of mitigations strategies for operational risks ranging from risky business practices to employee fraud (Ferguson 2012).

Therefore, analytically sophisticated companies put analytics to use in the widest possible range of decisions, especially to guide future strategies and day-to-day operations (LaValle, Lesser et al. 2011). Most of these applications are made possible by the use of ubiquitous sensors. Sensor technology, in the form of mobile, aerial or remote devices, as well as wireless internet networks, software logs, cameras, microphones, radio-frequency identification readers and so on, substantially contribute to the expansion of datasets (Philip Chen and Zhang 2014). Sensors can be used to monitor both objects and subjects and statistical models can be applied to make inferences on their behaviour (Gandy 2012). The concept of **Dataveillance As Targeted Analytics** (DVEIL) can then be applied to identify those cases in which the organisation collects data to monitor individuals' activities, analyses personal data to foresee and influence people's behaviour and relies on profiling to target valuable users or to personalise offers. These practices, despite being potentially controversial from a privacy perspective, can be seen as an important part of an enterprise's competitive strategy, especially in areas such as marketing (Ashworth and Free 2006). Analytically sophisticated enterprises are expected to manage a large variety of data and to rely on analytics to achieve different objectives in various functional units; it is thus reasonable to expect that concepts such as analytical sophistication, data pool variety and use of analytics across business functions will be positively associated with each other. These insights lead the researcher to formulate the following propositions and hypotheses summarised in figure 9.

> ***Proposition C.1****: The more analytically sophisticated (SOPH) organisations are, the more likely it will be that they employ analytics across business functions (FUNC) and process a large variety of data (DPOOL).*
>
> ***Hypotheses:***
>
> - *H-C.11 - SOPH will be positively associated with FUNCT.*
> - *H-C.12 - FUNC will be positively associated with DPOOL.*

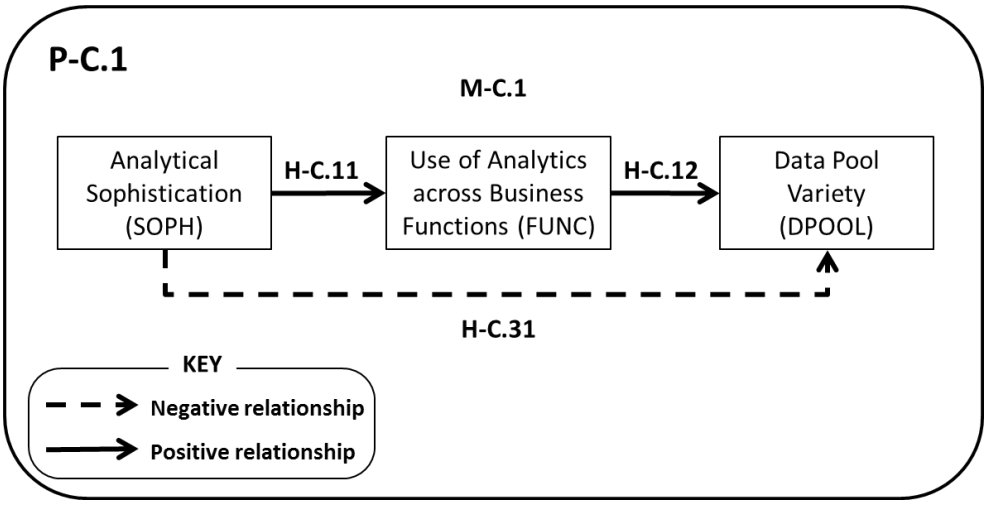
The fact that the enterprise relies on analytics across business functions might also drive the demand for data integration and increase the amount and variety of data processed by the

organisation. These considerations lead the researcher to formulate the following additional hypothesis.

Mediation effect:

- *M-C.1 - FUNC will mediate the relationship between SOPH and DPOOL.*

Figure 9. Proposition C.1 and corresponding hypotheses



Because of their characteristics, analytically sophisticated enterprises are also more likely to rely on target analytics, whose programs also demand the integration of different streams of data. These insights lead the researcher to formulate the following propositions and hypotheses, as summarised in figure 10.

Proposition C.2*: The more analytically sophisticated (SOPH) organisations are, the more likely it will be that they employ targeted analytics (DVEIL) on a large variety of data (DPOOL).*

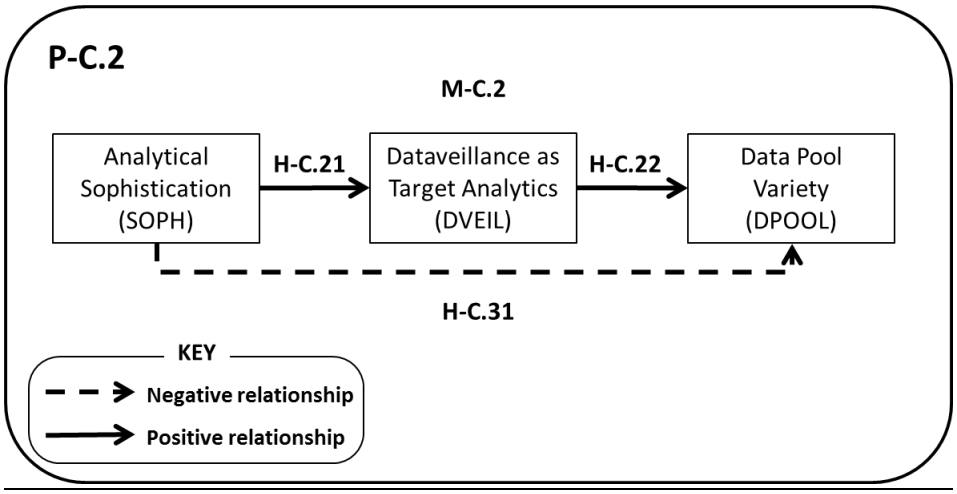
Hypotheses:

- *H-C.21 – SOPH will be positively associated with DVEIL.*
- *H-C.22 – DVEIL will be positively associated with DPOOL.*

Mediation effect:

- *M-C.2 - DVEIL will mediate the relationship between SOPH and DPOOL.*

Figure 10. Proposition C.2 and corresponding hypotheses



4.6.1 Analytical sophistication and data protection

Companies may extract value from personal information in a number of ways. Customers' data are the most common type of data nowadays analysed by private companies. Marketing scholars have widely recognized the key role played by customer data in boosting advertisement and customer-relationship marketing (Mouncey 2010). Data sharing can improve firms' marketing capabilities through targeted and online advertising, consumer recommendation systems and profit-enhancing price discrimination policies (Varian 1985, Acquisti and Varian 2005). Companies can minimize inventory risks and maximize returns on marketing investment given that the processing of customer data boost companies' ability to predict aggregate trends, such as variations in consumer demand. Furthermore, aggregated customer data can be used to forecast future demand and emerging trends.

Organisations have started recognising that corporate access to personal information must be balanced against a legitimate right of consumers to privacy (Culnan and Bies 2003). As privacy breaches result from poor organisational privacy practices (Chan, Culnan et al. 2005), privacy protection should be seen by business as an opportunity rather than as a threat. Corporations

can enhance their privacy programs by moving beyond merely complying with laws and other regulations and creating a culture of integrity that combines a concern for the law with an emphasis on managerial responsibility for the firm's organisational privacy behaviours (Culnan and Williams 2009).

The protection of information privacy may also be compatible with competitiveness and big data analytics as explained by Professor Thomas Davenport: "Stage 5 firms [Analytical Competitors] follow the Hippocratic oath of information privacy: above all, they do not harm. They have well-defined privacy policies […]. They don't break the privacy laws […]. They don't lose information […]. They don't sell or give away information without the permission of the customer or employee" (Davenport, Harris et al. 2010: p. 34). Companies need to have a data governance or information management process in place to ensure the data is clean as the value of data for decision-making purposes will be jeopardized if the data is not accurate or timely (SAS 2014).

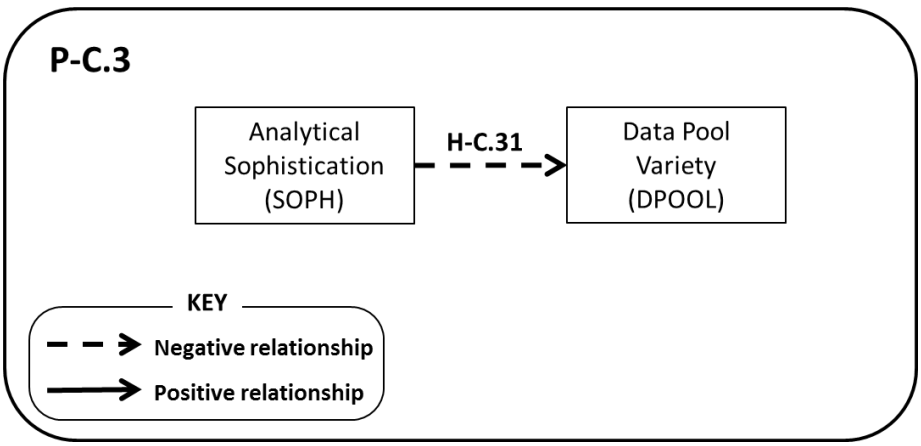
To be successful with big data, organisations need to develop processes and policies that accommodate new protocols for managing data privacy and security (Roski, Bo-Linn et al. 2014). Challenges in Big data analysis include data inconsistence and incompleteness, scalability, timeliness and data security; additional difficulties lie in data capture, storage, searching, sharing, analysis, and visualization (Philip Chen and Zhang 2014). Issues of data quality, privacy and security, and effectiveness of analysis are critical, for instance, in healthcare informatics (Kambatla, Kollias et al. 2014). A number of data pre-processing techniques, including data cleaning, data integration, data transformation and date reduction, have to be applied to remove noise and correct inconsistencies in noisy data.

A survey of 2,037 professionals and interviews with more than 30 executives reveals companies feel under pressure to improve their analytics capabilities, but there are signs that they may be getting overwhelmed by data management challenges (Kiron, Prentice et al. 2014). Streaming data cannot be stored in traditional ways and need to be analysed in real time (Anderson and Hardin 2014). Cleaning data and using only the data needed to solve specific business problems

are fundamental steps which guarantee to lower the risks of working with outdated, noisy data which may produce inaccurate results (Zicari 2013). Although big data has changed the way we capture and store data, including data storage device, data storage architecture, data access mechanism, more storage mediums and higher I/O speed is still required. Traditional enterprise storage architectures, such as direct-attached storage (DAS), network-attached storage (NAS), and storage area network (SAN), have severe drawbacks and limitations when it comes to large-scale distributed systems (Philip Chen and Zhang 2014). On the other side, network bandwidth capacity is the bottleneck in cloud-based systems. In other words, indiscriminate data accumulation can generate problems to analytically sophisticated enterprises which could decide to adopt measures to reduce the amount or variety of data collected in order to ensure data quality. These insights lead the researcher to formulate the following proposition and hypotheses, as summarised in figure 11.

> ***Proposition C.3****: The more analytically sophisticated organisations are, the less likely it will be that they collect and store indiscriminately a large variety of data.*
>
> ***Hypothesis:***
>
> - *H-C.31 - SOPH will be negatively associated with DPOOL*

Figure 11. Proposition C.3 and corresponding hypotheses

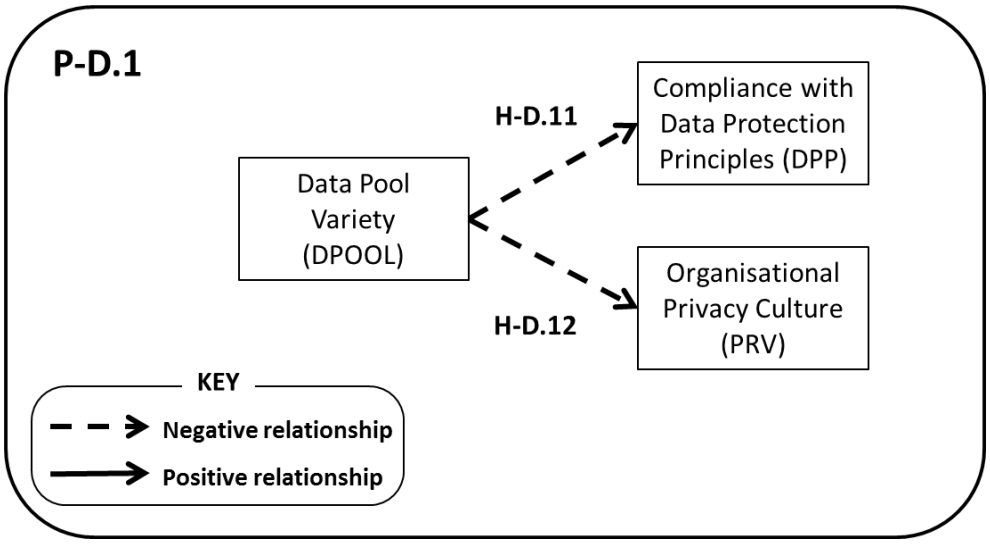


For organisations operating online collecting information regarding internet users' characteristics and behaviour has become extremely easy. The same architecture of the Internet allows multiple

parties to collect data and compile identifying information to elaborate detailed user profiles (Tene 2010). Online users do not receive any notice of being tracked while they are surfing the Internet (Datta, Tschantz et al. 2015). For third-party organisations which gather information on unaware internet users' browsing activity on various unrelated websites and sell this information to facilitate the targeting of advertisement by grouping people according to certain parameters, the chances of complying with European data protection laws could be limited. Inadequate enforcement capacity and fragmentation of regulatory requirements across jurisdictions could also reduce the effectiveness of legal actions. As pointed out by Mark Andrejevic, "[i]n the petabyte era, […] collecting information about everyone becomes not just a technological temptation, but an operational necessity" (Andrejevic 2009: p. 322). Although technological solutions, such as services like 'DoNotTrackMe', have been developed to help users stop third parties collecting information about them (Abine 2015), tracking technologies are also in constant evolution (Lecuyer, Spahn et al. 2015); information privacy represents a challenge for data brokers and other data harvesting companies, which are secretive about their data sources and on revealing the identity of their business clients (Otto, Antón et al. 2006, US-Senate 2013). These insights lead the researcher to formulate the following proposition and hypotheses, as summarised in figure 12.

> ***Proposition D.1****: The more organisations process and analyse a large variety of data (DPOOL), the less likely it will be that they develop a privacy culture (PRV) and comply with data protection principles (DPP).*
>
> ***Hypotheses:***
>
> - *H-D.11 - DPOOL will be negatively associated with DPP.*
> - *H-D.12 - DPOOL will be negatively associated with PRV.*

Figure 12. Proposition D.1 and corresponding hypotheses



4.6.2 Targeted analytics as a form of dataveillance

The availability of data about all aspects of customers' lives, raises concerns about their autonomy under marketers' gaze (Lace 2005). Techniques such as 'nudging' conceived to exploit cognitive biases in order to persuade people to behave in a certain way are also gaining popularity (Calo 2013, Cavoukian, Stewart et al. 2014). These apparently innocuous techniques have a lot in common with traditional surveillance tools, which were designed to persuade people to behave in a certain way simply by triggering certain psychological reactions. A typical example is the 'chilling effect' surveillance may produce on the monitored subject (Askin 1972, Hughes 2012).

It is common practice for analytical competitors to rely on sophisticated experiments, in order to measure the overall impact, or 'lift', of different intervention strategies (Davenport 2006). Business experiments apply the scientific method to determine whether a particular business intervention is effective or not (Davenport 2007). Companies also rely on experiments to validate their knowledge and take so-called 'fact-based decisions' (Harris, Morison et al. 2010). Online platforms rely heavily on experiments as an innovation and performance assessment tool.

An aspect which raises special concerns refers to the way dataveillance is changing people's behaviour and reducing people's choice space by means of experimental methods conceived to

manipulate individual actions (Degli Esposti 2014). According to the business literature, the proliferation of predictive models to forecast people's future behaviour helps companies increase sales by anticipating customers' desires (Davenport and Harris 2009). For example Facebook Likes can be used to accurately predict highly sensitive personal attributes such as sexual orientation, ethnicity, religious/political views, degree of happiness, addictive substance consumption, parental separation, and so on (Kosinski, Stillwell et al. 2013, Kshetri 2014).

Specific methodologies help analysts anticipate people's behaviour. Prescriptive modelling allows analysts to include information on preferred outcomes and evaluate alternative course of actions against potential results. Predictive modelling is used, for example, to identify the most 'valuable' customer, plus those with the greatest 'profit potential' and the ones most likely to cancel their accounts.

One of the most controversial areas in terms of privacy implications is certainly behavioural targeting. Dataveillance in the form of behavioural targeting can be used to influence people's behaviour in subtle ways (Degli Esposti 2014). Similarly to other technological innovation, big data bring risks alongside opportunities for both individuals and organisations. In particular, it gives organisations pervasive and massive surveillance capabilities at a very low cost. The possibility of tracking individual activities both online and offline has serious privacy implications. The following aspects are considered especially controversial (Nunan and Di Domenico 2013): (a) the combination of previously fragmented data or dispersed datasets; (b) the risk of data being stolen or accessed by unauthorised malicious people; (c) the proliferation of automated decision-making algorithms replacing human judgement; (d) the automatic collection and retention of all data going through a network due to decreasing storage costs; (e) the impossibility of foreseeing how data will be used in the future when more advanced analytical tools will become available.

Surveillance studies scholars tend to consider the concept of privacy, and the policies it generates, fundamentally inadequate (Stalder 2002).The privacy regime has been criticised for being "too narrow, too based on liberal assumptions, too implicated in rights-based theory and discourse,

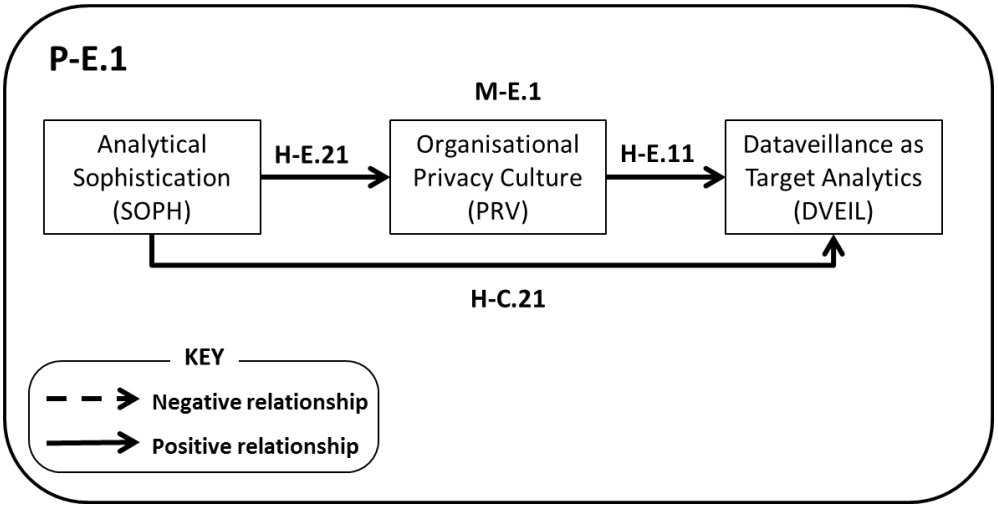
insufficiently sensitive to the social sorting and discriminatory aspects of surveillance, culturally relative, overly embroiled in spatial metaphors about 'invasion' and 'intrusion', and ultimately practically ineffective" (Bennett 2011: p. 485). The privacy regime is considered insufficient in addressing the collective effects produced by digital surveillance (Gilliom 2001, Gilliom 2011). In the case of facing very distressing experiences, such as a house search by security forces, the concept and the legal regime based on the concept of 'privacy' are inadequate in conveying all the senses of anger, frustration, disbelief, and absurdity of the experience (Gilliom 2001, Gilliom 2011).

> [P]rivacy is a weak argument in the face of overwhelming arguments for public safety, drug-control, accountability and welfare fraud control; privacy has a cultural weakness as a NIMBY like, me-first sort of value; privacy has, for the most part, become a procedural order, not a substantive guarantee (Gilliom 2011: p. 503).

The idea of 'privacy' seems also to be inadequate in constraining and limiting dataveillance at organisational level: the privacy discourse helps people and institutions interpret 'social sorting' as an additional personalised service and as a 'positive' form of social discrimination based on institutional control over personal information (Stalder 2011). As also explained in previous sections, the development of the organisational privacy culture goes hand in hand with the adoption of sophisticated analytical tools. These insights lead the researcher to formulate the following propositions and hypotheses, as summarised in figure 13 and 14.

> ***Proposition E.1****: The more organisations are analytically sophisticated (SOPH), the more likely it will be that they develop an organisational privacy culture (PRV) compatible with the use of targeted analytics (DVEIL).*
>
> ***Hypotheses:***
>
> - *H-E.11 – DVEIL will be positively associated with PRV.*
>
> ***Mediation effect:***
>
> - *D-E.1 – PRV will mediate the relationship between SOPH and DVEIL.*

Figure 13. Proposition E.1 and corresponding hypotheses



Proposition E.2*: The more organisations are analytically sophisticated (SOPH) and rely on analytics across business functions (FUNC), the more likely it will be that they develop an organisational privacy culture (PRV).*

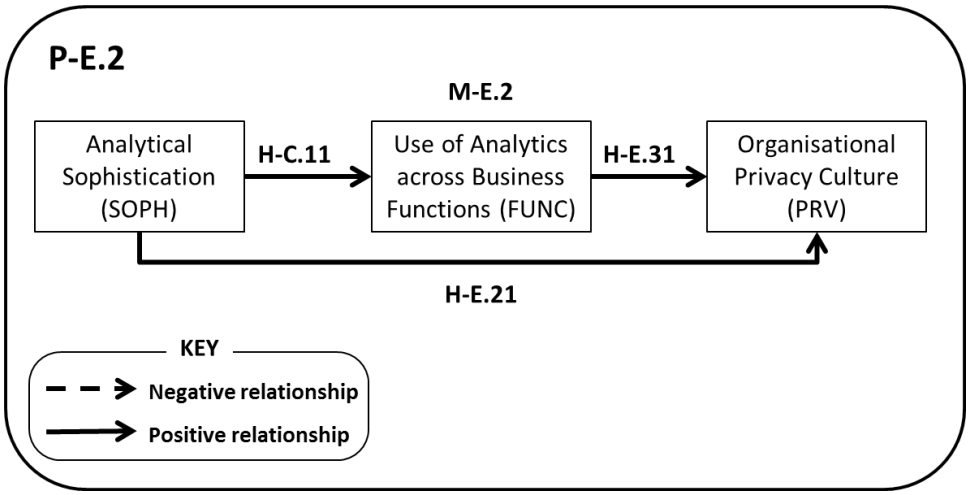
Hypotheses:

- *H-E.21 - SOPH will be positively associated with PRV.*

Mediation effect:

- *M-E.2 - FUNC will mediate the relationship between SOPH and PRV.*

Figure 14. Proposition E.2 and corresponding hypotheses



Proposition E.3*: The more organisations rely on analytics across business functions, the more likely it will be that they develop an organisational privacy culture and use targeted analytics in a privacy-respectful way.*

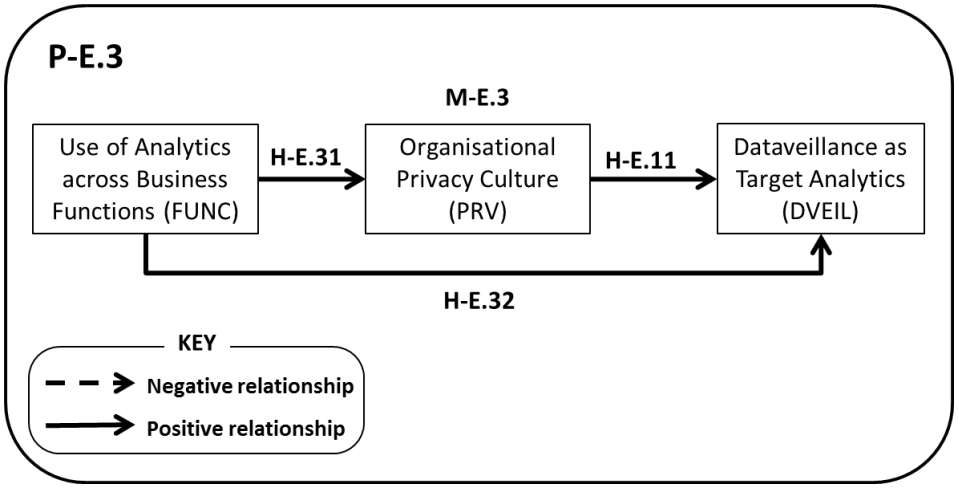
Hypotheses:

- *H-E.31 - FUNC will be positively associated with PRV.*
- *H-E.32 - FUNC will be positively associated with DVEIL.*

Mediation effect:

- *M-E.3 - PRV will mediate the relationship between FUNC and DVEIL.*

Figure 15. Proposition E.3 and corresponding hypotheses



4.7 Motivations behind the allocation of resources to data protection initiatives

Organisations can interpret privacy protection either as a risk or as an opportunity (Greenaway and Chan 2013). Companies which view privacy protection as a risk to the firm tend to strive to minimise the likelihood of data breach and of scrutiny by privacy regulators. In contrast, companies which see privacy protection as an opportunity are keen on investing to improve customer relationships in order to meet customers' information privacy expectations. Heng and co-authors define “institutional privacy assurance as the interventions that a particular company makes to ensure consumers that efforts have been devoted to protect personal information”

(Heng, Dinev et al. 2011: p. 805). Typical interventions are company privacy policy and industry self-regulation.

As the primary source of data within organisations comes from consumers, marketing is the function whose privacy risks become more apparent. Although privacy was already a significant marketing issue in the 1990s (Jones 1991), noteworthy data breaches, such as the Sony case, have raised business awareness of the reputational, legal and economic consequences of information security corporate scandals (Kieke 2014). When personal data are accessed, or stolen, by unauthorised agents organisations need to revise their information security policies and procedures and to give explanations to the public. For this reason, the need to ensure the protection of customer information and to comply with regulation may contribute to increase organisations' security expenditures (BIS 2014). Strong enforcement actions by regulators can also help transform data security into a priority within specific business sectors – like in the case of the protection of electronic patients' records in the U.S. since the enforcement of the Health Insurance Portability and Accountability Act, known as HIPAA (Kieke 2014). However, organisations may respond in a heterogeneous way to institutional pressures related to information security by making different levels of investment (Cavusoglu, Cavusoglu et al. 2015). The way organisations assess their security needs, among other factors, play also a role in shaping information security decisions. Because of the importance if these topics, an exploration of potential reasons behind information security investment decisions has also been included in this study.

4.8 Summary

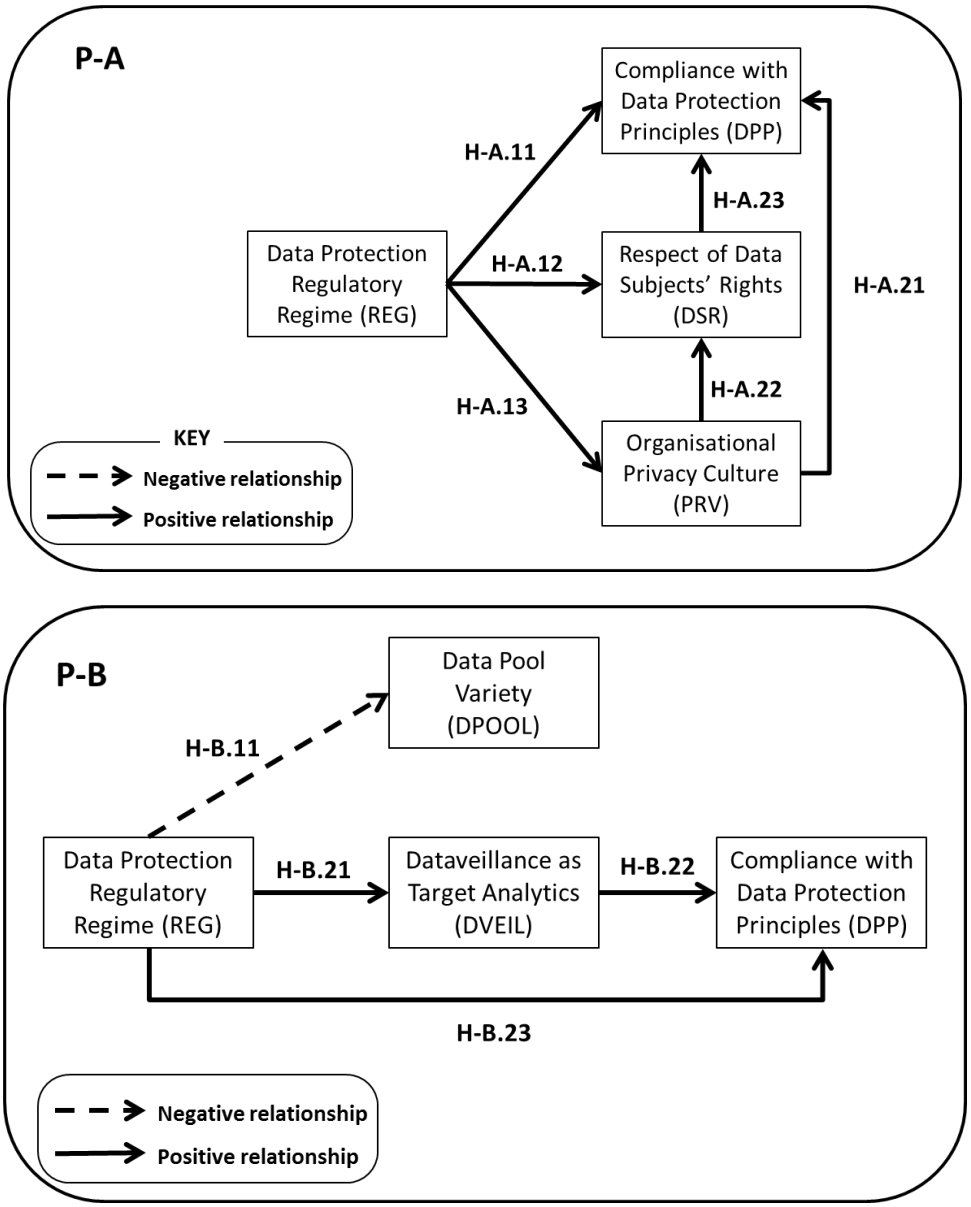
While Chapter Two and Three have presented the context of this study, within this chapter a set of propositions have been identified in the attempt to answer the following research questions: (1) how does the data protection regulatory regime influence enterprise data protection and data management decisions? (2) How does the level of analytical sophistication, achieved by an organisation, influence enterprise data protection and data management decisions?

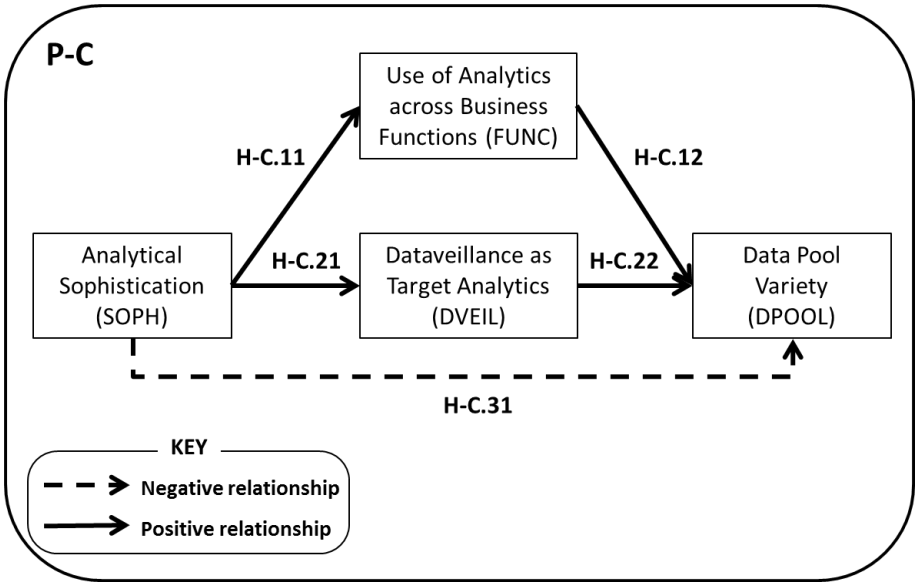
There are in total five groups of propositions, each of them identified by a letter. Within each group individual propositions and hypotheses are identified by the same letter and a number. Propositions A and B, and their corresponding hypotheses, explore the effect of the data protection regulatory regime on the degree of compliance with data protection principles, on the level of respect of data subjects' rights, and on the creation of the organisational privacy culture. The effect of the data protection regulatory regime on the use of targeted analytics and massive data collection is explored in proposition B. As European data protection law is expected to prevent the indiscriminate accumulation of data, a negative relationship is supposed to exist between a strict and reliable regulation and data collection. These models are displayed in figure 16.

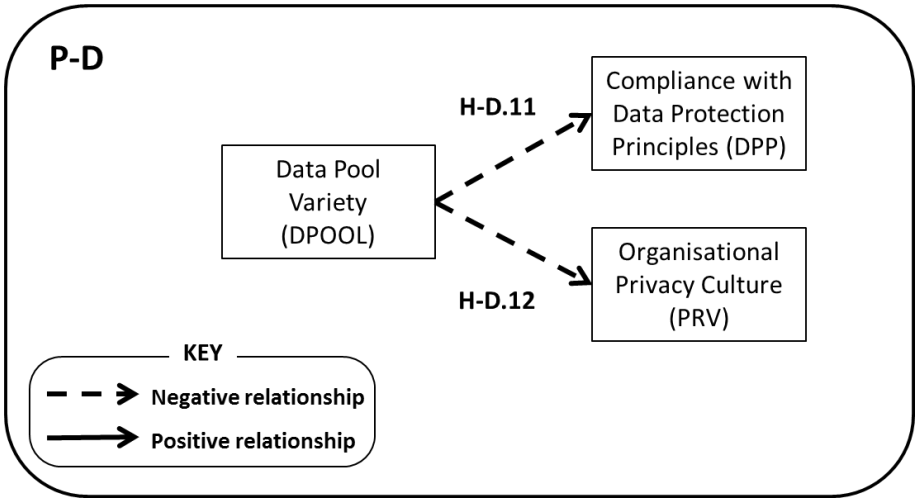
On the other hand, propositions C, D, and E, and their corresponding hypotheses, are meant to answer the second research question, which explores the effects of analytical sophistication on the corporate data privacy environment. These models are also displayed in figure 16. According to previous studies, and as expressed in proposition C, analytically sophisticated organisations process a large amount of data in different formats and rely on analytics across business functions to achieve their objectives; they also employ targeted analytics to personalise offers and influence customers' behaviour. Analytically sophisticated organisations pay also attention to data quality; they try not to harvest data indiscriminately and reduce the risk of creating unnecessary noise in the data or the risk of suffering a data breach. For this reason, in proposition C, it has been identified a negative relationship between analytical sophistication and data pool variety. Since a considerable amount of data processed by organisations are personal data, proposition E identifies a self-reinforcing mechanism between relying on analytics and establishing an internal data privacy culture. Nonetheless, the more organisations collect and process personal data, the more challenging it becomes to comply with data controllers' obligations. Thus, proposition D points out the main point of friction between the logic of big data and the logic of data protection, which refers to data accumulation. Massive data harvesting performed by organisations is

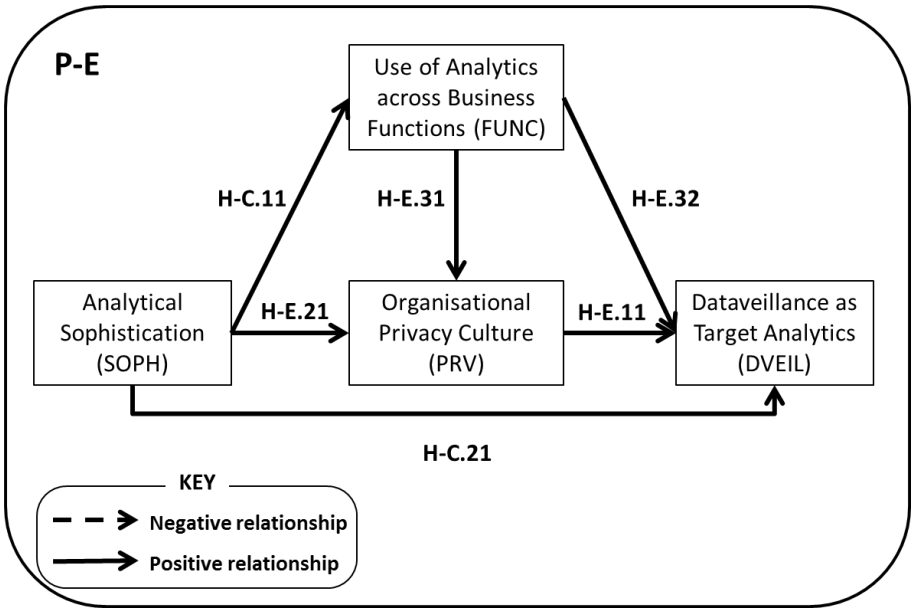
expected to lower their chances to comply with data protection principles and establish an internal privacy culture.

Figure 16. Propositions A and B answering question one and propositions C, D, and E answering question two




P-C
Use of Analytics across Business Functions (FUNC)
H-C.11
H-C.12
Analytical Sophistication (SOPH)
H-C.21
Dataveillance as Target Analytics (DVEIL)
H-C.22
Data Pool Variety (DPOOL)
H-C.31
KEY
Negative relationship
Positive relationship


P-D
H-D.11
Compliance with Data Protection Principles (DPP)
Data Pool Variety (DPOOL)
Organisational Privacy Culture (PRV)
H-D.12
KEY
Negative relationship
Positive relationship


P-E
Use of Analytics across Business Functions (FUNC)
H-C.11
H-E.31
H-E.32
Analytical Sophistication (SOPH)
H-E.21
Organisational Privacy Culture (PRV)
H-E.11
Dataveillance as Target Analytics (DVEIL)
H-C.21
KEY
Negative relationship
Positive relationship

Thus, in explaining the level of compliance with data protection principles, a relevant factor is the *data protection regulatory regime* enforced in the national context where the organisation operates. Another relevant element is the *internal privacy culture* the organisation was able to establish. According to considerations based on other studies and reported in previous chapters, the data protection regulatory regime is expected to limit the indiscriminate harvesting of data, while also fostering a privacy-conscious use of targeted analytics.

Moving to the big data side, attention is paid to the effect that *analytical sophistication* may have on the likelihood of adopting *targeted analytics*, and analytics across business function to pursue organisational objectives. The effects of *targeted analytics*, *analytical sophistication*, and the *use of analytics across business functions* on *data pool variety* and accumulation, aspects with potentially serious privacy implications, have also been assessed. The information management system an organisation has developed to become analytically sophisticated supports the adoption of analytics across business functions, the use of targeted analytics and the proliferation of activities to raise privacy awareness. In contrast, the accumulation of large amounts of data in different formats to be used for achieving various objectives may create friction with respect to data subjects' rights and data protection principles. Nonetheless, compromises can be found between considering privacy a core organisational value and using targeted analytics to achieve organisational objectives. Developing an internal privacy culture is also part of the path toward analytical sophistication.

The overall model with all hypotheses is presented in figure 17 and summarised in table 10.

Figure 17. Theoretical model with all hypotheses

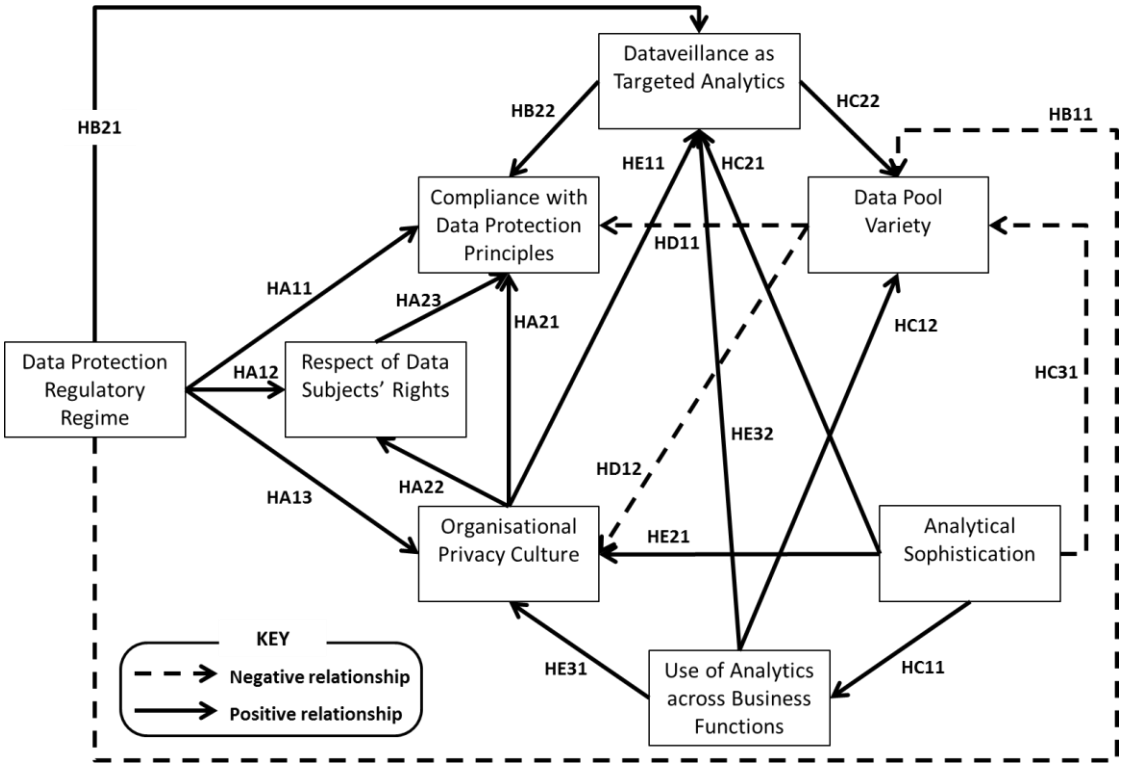


Table 10. Summary: List of all propositions with hypotheses

Proposition	Hp	DV ⇦ IV	Hypothesis	Indirect effect
Proposition A.1: *The less permissive and more reliable the data protection regulatory regime, the more likely it will be that organisations develop an internal privacy culture, respect data subjects' rights and comply with data protection principles.*	***H-A.11***	DPP ⇦ REG	REG will be positively associated with DPP	
	H-A.12	DSR ⇦ REG	REG will be positively associated with DSR	
	H-A.13	PRV ⇦ REG	REG will be positively associated with DSR	
Proposition A.2*: The more organisations foster their internal privacy cultures, the more likely it will be that they respect data subjects' rights and that comply with data protection principles.*	***H-A.21***	DPP ⇦ PRV	PRV will be positively associated with DPP	***M-A.2*** DSR will mediate the relationship between PRV and DPP
	H-A.22	DSR ⇦ PRV	PRV will be positively associated with DSR	
	H-A.23	DPP ⇦ DSR	DSR will be positively associated with DPP	

Proposition B.1*: The less permissive and more reliable the data protection regulatory regime, the less likely it will be that organisations indiscriminately collect and analyse a large variety of data.*	***H-B.11***	DPOOL ⊂ REG	REG will be negatively associated with DPOOL	
Proposition B.2*: The less permissive and more reliable the data protection regulatory regime, the more likely it will be that organisations use targeted analytics in a way compliant with data protection principles.*	***H-B.21***	DVEIL ⇦ REG	REG will be positively associated with DVEIL	***M-B.2*** DVEIL will mediate the relationship between REG and DPP
	H-B.22	DPP ⇦ DVEIL	DVEIL will be positively associated with DPP	

Proposition C.1*: The more analytically sophisticated organisations are, the more likely it will be that they employ analytics across business functions and process a large variety of data.*	***H-C.11***	FUNC ⇦ SOPH	SOPH will be positively associated with FUNC	***M-C.1*** FUNC will mediate the relationship between SOPH and DPOOL
	H-C.12	DPOOL ⇦ FUNC	FUNC will be positively associated with DPOOL	
Proposition C.2*: The more analytically sophisticated organisations are, the more likely it will be that they employ targeted analytics and process a large variety of data.*	***H-C.21***	DVEIL ⇦ SOPH	SOPH will be positively associated with DVEIL	***M-C.2*** DVEIL will mediate the relationship between SOPH and DPOOL
	H-C.22	DPOOL ⇦ DVEIL	DVEIL will be positively associated with DPOOL	
Proposition C.3*: The more analytically sophisticated organisations are, the less likely it will be that they collect and store indiscriminately a large variety of data.*	***H-C.31***	DPOOL ⊂ SOPH	SOPH will be negatively associated with DPOOL	

Proposition	**Hp**	**DV ⇦ IV**	**Hypothesis**	**Indirect effect**
Proposition	**Hp**	**DV ⇦ IV**	**Hypothesis**	**Indirect effect**
Proposition D.1*: The more organisations process and analyse a large variety of data, the less likely it will be that they develop a privacy culture and comply with data protection principles.*	***H-D.11***	DPP ⊂ DPOOL	DPOOL will be negatively associated with DPP	
	H-D.12	PRV ⊂ DPOOL	DPOOL will be negatively associated with PRV	

Proposition E.1*: The more organisations are analytically sophisticated, the more likely it will be that they develop an organisational privacy culture compatible with the use of targeted analytics.*	***H-E.11***	DVEIL ⇦ PRV	DVEIL will be positively associated with PRV	***M-E.1*** PRV will mediate the relationship between SOPH and DVEIL
Proposition E.2*: The more organisations are analytically sophisticated and rely on analytics across business functions, the more likely it will be that they develop an organisational privacy culture.*	***H-E.21***	PRV ⇦ SOPH	SOPH will be positively associated with PRV	***M-E.2*** FUNC will mediate the relationship between SOPH and PRV
Proposition E.3*: The more organisations rely on analytics across business functions, the more likely it will be that they develop an organisational privacy culture and use targeted analytics in a privacy-respectful way.*	***H-E.31***	PRV ⇦ FUNCT	FUNC will be positively associated with PRV	***M-E.3*** PRV will mediate the relationship between FUNC and DVEIL
	H-E.32	DVEIL ⇦ FUNCT	FUNC will be positively associated with DVEIL	

KEY: Analytical Sophistication (SOPH); Dataveillance (DVEIL); Data Pool Variety (DPOOL); Data Protection Regulatory Regime (REG); Compliance with Data Controllers' Obligations (DPP); Respect of Data Subjects' Rights (DSR); Organisational Privacy Culture (PRV); Use of Analytics across Business Functions (FUNC).

4.9 Conclusions

This study hopes to contribute to the dearth of prior organisational-level privacy research, which has largely overlooked ethical issues or the personal harms often caused by privacy violations (Culnan and Williams 2009). Special attention is paid to the effects of (1) data protection legislation and (2) analytical sophistication on the corporate data privacy environment, which includes (a) the adoption of basic data protection principles and (b) the establishment of a privacy culture within the organisation. The term 'data protection principles' refers to the terminology adopted by European data protection legislation. Since this study is meant to shed light on the relationship between big data, analytics, and data protection, particular attention has been paid to the use made by organisation of targeted analytics. The use of analytics to target customers, or service users, is a particularly controversial area. The importance of paying attention to the ethical implications of relying on targeted analytics has been recognised in previous privacy studies, especially in the area of database marketing (Thomas and Maurer 1997, Milne 2000, Petty 2000). This study tries to overcome the limits of addressing the topic of data protection only from an information privacy perspective, by introducing the concept of dataveillance, proceeding from surveillance studies. By interpreting targeted analytics as a form of dataveillance, this study hopes to shed light on the complex relationship between the current privacy legal regime and technology trends like big data. It will also help test whether dataveillance and the current privacy regimes are compatible, as pointed out by surveillance scholars (Bennett 2011, Gilliom 2011).

The next chapter presents the methodology and the operationalisation of all constructs, followed by the analysis of the data in Chapter Six and a discussion of the implications of the study's results in Chapter Seven.

CHAPTER FIVE

Research Design and Methodology

5.1 Introduction

Methodological considerations about how the researcher has run this study are reported in this chapter, which is divided into two main parts: the development of the scales and the questionnaire; and instrument preliminary testing procedures. Initially this chapter presents the various steps the researcher has followed in constructing the data gathering instruments. The next chapter presents the data collection strategy and explains the set of statistical techniques the researcher has used to analyse the data and test the hypotheses. Topics such as participant recruitment strategy, sample selection, and non-response bias are also discussed in the next chapter.

5.2 Epistemological considerations

In framing this project, it is necessary to make explicit some general assumptions at the outset. This research maintains a realist ontological stance, but also recognises that 'scientific' evidence emerges from the interaction of the inquirer and the phenomenon inquired; it also is aware of the fallacy of theories, imprecision of measures, and need for the inquirer to be critical and reflective (Guba 1990). In particular, this project adopts scientific realism as an epistemological perspective (Chakravartty forthcoming). In contrast with antirealism perspectives, such as social constructivism or feminist approaches, scientific realism keeps the notion that the inherent order of things is 'mind-independent', that causation must be distinguished from correlation, and that research findings should not be generalized unless they can be replicated across samples, populations and research methods (Mir and Watson 2001). Taking a realist approach implies the adjustment of the research method to the object under study, and a clear recognition of the

underlying study's assumptions and limitations produced by the method used (Mingers 2004). In this respect, quantitative and qualitative methods should be combined to balance precision and richness of information about the causal mechanism underlying the object of inquiry and triangulation of research methods would always be advisable. While positivists tend to neglect all that part of reality that is not empirically quantifiable and to rely on contest-free experiments, realists consider both the observable and unobservable of the world as objects of potential inquiry and pay special attention to characteristics of the field where the research is undertaken (Hacking 1982). In this respect, within this project qualitative interviews have been carried out to improve the quality of questionnaire items, quantitative findings have been contrasted with practitioners' experiences as reported in grey literature or previous studies, and various statistical techniques have been used to assess the robustness of certain results. As such the major application of this theory in research is explaining complex social events, detecting underlying structures that generate, or do not generate, particular patterns of events, and ruling out any other potential explanations.

Realist approaches have been widely adopted in business studies. In information systems research, it was considered a fruitful perspective for overcoming inconsistencies between theoretical assumptions and actual research practices adopted (Mingers 2004, Smith 2006) and for justifying the adoption of case study research (Fox 2009, Easton 2010). In accounting, it has been interpreted as a way to retain elements of scientific rigour, and yet acknowledge the value of richness and context, as well as the importance of generalizability (Bisman 2010). In business ethics, these approaches are considered suitable for understanding both socially beneficial and harmful corporate practices and for allowing logics, which are not strictly financial, to emerge as important explanatory factors underpinning organisational dynamics (Wry 2009).

5.3 Research Design

This study is meant to shed light on the antecedents of organisational compliance with data protection principles by exploring the effects of the data protection regulatory regime and the

level of analytical sophistication on organisational information management decisions. In pursuing this goal, the study draws insights from previous organisational privacy studies (Milberg, Smith et al. 2000) and on the analytical competitor literature (Davenport and Dyché 2013). In contrast with previous studies on analytical competitors' characteristics (Davenport and Harris 2007, Davenport 2014), which have mostly relied on qualitative data, this study relies on survey data and quantitative statistical techniques, following the tradition of previous organisational privacy studies (Milberg, Smith et al. 2000). Despite relying on quantitative methods, the nature of the study is exploratory. This research in fact represents the first attempt to investigate the relationship between data protection and big data analytics at organisational level. In doing so, it draws insights mostly from qualitative studies in the area of surveillance studies (Ball 2010), privacy studies (Bamberger and Mulligan 2011) and the literature on analytical competitors (Davenport 2014). Although the theoretical model with propositions and hypotheses presented in chapter four is based on the knowledge accumulated in these studies, several aspects investigated here have not been investigated before. As a result, the same use of path analysis made in chapter six to test hypotheses has to be interpreted most as an exploratory and theory-building exercise than as a confirmatory exercise. On the other hand, the collection of quantitative data allowed the researcher to increase the number of organisations with different characteristics represented in the study and to make comparisons in a more systematic way.

Regarding the methodology, an electronic survey was chosen as the most suitable research method for gathering relevant information about organisations' information management practices related to both data protection and big data analytics. As it is common in organisational research, key informants have provided information on the properties and characteristics of their organisations (Bagozzi, Yi et al. 1991). In order to increase the chances to reach the audience of business professionals who were knowledgeable about their enterprises' information management procedures, a 'relevant media' strategy has been adopted. To reach those experts capable of providing information on their organisations' information management internal procedures, specific online media channels were identified and used to advertise the study.

Recognised online publications in the area of data protection laws or professional groups of IT experts or other professional associations contributed to publicise the study. This 'relevant media' strategy contributed to lower the risks of facing a low response rate.

5.4 Survey design

In order to collect data about companies' data protection and management practices, the project has taken the form of an online survey. This method has many advantages: it gives respondents the chance to reply when it is more convenient for them; once the questionnaire is filled in, it can be returned immediately to the researcher; interviewer's or researcher's bias is limited (Miyazaki and Taylor 2008); and using this method is relatively low cost. Relying on an electronic questionnaire creates also some risks, such as low response rate, and implies certain limitations related to the quality of responses. For instance, having a long survey was unfeasible and questions had to be grouped and organised to produce a concise yet complete questionnaire.

Internet-based surveys can suffer from responses from individuals outside the population of interest or multiple responses from a single individual, elements which would both lead to biased results (Schillewaert, Langerak et al. 1998). To reduce these risks, various communication methods were used to elicit contact and cooperation with respondents. To build trust and cooperation, information about the output of the survey, debriefing procedures and privacy safeguards were also included in the communication (Snijkers, Haraldsen et al. 2013). A report with study results was also made available, as part of the debriefing phase, in order to foster participation. As quality of response could not be assessed while interviewees were responding, some control items and screen-out questions were also added to ensure the reliability of responses and avoid the same person filling in the questionnaire more than once.

Survey items were continuously revised to avoid bias with wording. Ambiguous, complex, or double-barrelled questions which used vague or uncommon words were deleted. A statement explaining the objective of the study and offering a definition of few relevant concepts were also provided at the beginning of the survey to give some context and avoid misinterpretation. Since

in self-administered questionnaires, horizontal vs. vertical format of the response choices, or left vs. right side alignment of the response choices, can affect the answers (Choi and Pak 2005), the same format was consistently used across survey items. On one-to-seven scales positive statements always appeared on the left hand side of the scale and a horizontal format was consistently used. A vertical format of the response choice was adopted in the case of multiple-answer items. Randomisation of survey items in cases where the ordering of the options was irrelevant was also introduced to avoid the risks that options at the beginning of the list were selected more often than those at the end of the list. To avoid missing or overlapping intervals in response choices which could cause confusion, all questions offering interval categories were revised to ensure categories were mutually exclusive (Choi and Pak 2005). Open questions, 'do not know' options and middle-point scales were introduced to cope with the constraints and lack of flexibility typical of questionnaires and to limit the risks of providing too few categories and forcing respondents to choose among limited options.

5.4.1 Effects of relying on online surveys on sampling characteristics

The 'relevant media' survey administration strategy produced a sample of usable surveys which has to be considered a 'convenience' sample, as it consists of a group of volunteers interested in the topic treated in the study. This type of non-probability samples has several limitations. First of all, the probability of a subject being selected cannot be computed because of the lack of a complete sampling frame. This is a well-known problem in internet research, characterised by the unavailability of a comprehensive list of e-mail addresses of the internet population (Taylor 1999) and limited coverage (Velu and Naidu 2009). A convenience sample can thus lead to the under-representation or over-representation of particular groups within the sample: the opinions of people with extreme views, either positive or negative, can be overemphasised. The units selected for inclusion in the sample are usually simply the easiest to access, which means that they do not represent accurately the characteristics of the entire population. A random sample, in contrast, adheres to the following criteria: (a) each subject in the population has an equal

likelihood of being selected as a member of the sample; (b) the selection of each subject is independent of the selection of all other subjects in the population; and (c) for a specified sample size, every possible sample that can be derived from the population has an equal likelihood of occurring (Pfeffermann and Rao 2009).

Another problem of dealing with a convenience sample is being unable to compute survey response rates. In this study, for instance, the researcher was unable to obtain information on the number of people who received information about the study. In general, business surveys tend to produce unsatisfying response rates because of lack of time, knowledge, and availability (Rasmussen and Thimm 2009). Practitioners nowadays are also flooded with questionnaires which often are considered not relevant; because of fatigue employees may refuse to respond to non-essential questionnaires or it can be a company policy not to complete surveys (Weiner & Dalessio, 2006). In sum, the two principal reasons for not responding are failure to deliver the questionnaires to the target population, because of problems with respondents' contact information, and the reluctance of people to respond (Baruch and Holtom 2008).

Therefore, the main disadvantage of relying on the 'relevant media' strategy has been the generation of a non-random sample; something fairly common in survey design studies where "[a] great deal of research is based on the use of nonprobability samples" (Sheskin 2003: p. 87). To overcome, or at least address, the limitations previously mentioned, the next section provides evidence that the survey distribution channel has not created any major source of bias. Checking for unexpected sources of variability due to the sample selection procedure is certainly an advisable procedure especially when a non-probability sample is applied without a complete sample frame (Hui-Chih and Her-Sen 2010). Since it was not possible to compute the survey response rate, the researcher decided to focus on the survey completion rate as a way to explore potential sources of non-response bias. This and related issues are discussed in the next section.

5.4.2 Research context

Since the study investigates the degree of compliance with data protection principles stated in the 1995 EU Data Protection Directive, the natural context of this study was meant to be the European Union. Nonetheless, the distribution channels selected were able to reach professionals working in other countries, especially the US. As it was impossible to reach professionals in all EU countries, the researcher has decided to focus her attention on the United Kingdom and Spain, as they represent cases where the difference between Data Protection Authorities' powers and resources are more evident. The Spanish Data Protection Authority is famous for the serious fines it can issue to companies, while the British Information Commissioner's Office (ICO) has started enjoying the power to impose more serious sanctions only in recent times. To increase the participation of British and Spanish professionals the invitation to participate in the survey was sent through the ICO's newsletter and to the members of the British Computer Society. Similarly, the Spanish Society of Privacy Professionals (APEP) and the Spanish Society of IT Experts (ATI) also invited their members to fill in the electronic survey.

5.5 Data collection strategy

Between January and April 2014, 442 professionals participated in the Big Data Protection Study by filling in an online survey. The UK Information Commissioner's Office (ICO) published two posts in its e-newsletter advertising the study (ICO 2013, ICO 2014). Several specialised media, such as the magazine *Privacy Laws and Business*, the Operational Research Society's magazine *Inside O.R.*, and the online blog IAPP's *Privacy Perspectives*, published articles about the study. The British Computer Society Effective Leadership in IT (ELITE) Group invited its members to participate. The *European Data Protection Supervisor* (EDPS), the *Spanish Association of Privacy Professionals* (APEP), and the *Spanish Association of IT professionals* (ATI) also contributed by distributing invitations to participate in the study.

Online media, professionals groups and associations were identified as suitable channels to reach the target audience and start the recruitment process. This 'relevant media' strategy was adopted

as a second-best solution after attempting to use a mailing list database acquired for this purpose. The *UK Business Email List Database*, bought on www.specialdatabases.com, contained 380,573 individual records of professionals working in the UK. The researcher partitioned the database into categories of interest based on people's job titles – i.e. marketing experts (10,584), IT professionals (8,065), and chief executive officers (9,353) – and extracted a random sample from each group. An invitation letter with the link to participate in the study and a communication opt-out option was sent on September 2013 to 15% of people from each group – 1,403 CEOs; 1,585 marketing professionals; and 1,209 IT professionals. The fact that emails were sent through the Qualtrics server and the presence of outdated email addresses increased dramatically the likelihood of the email been detected and categorised as 'spam'. A person wrote to the researcher asking if the mail was spam. In total only six people started the survey and only three of them completed it.

To overcome the problem of dealing with outdated emails, a 'relevant media' recruitment strategy was subsequently adopted. The 'relevant media' recruitment strategy is a type of recruitment strategy which identifies and uses journals, blogs, professional groups or other channels visited by members of the group under study, in this case privacy and legal officers and IT managers. The main advantage of adopting a 'relevant media' strategy was to be able to reach target groups, such as information systems and privacy experts. By publishing articles on specialised online media, or by inviting special groups to participate in the study, this strategy helped the researcher to increase the response rate at a very low cost. Since the research focused on both data protection and information systems aspects of the big data phenomenon, in order to ensure coverage of both the population of privacy professionals and IT professionals, alternative channels were used to reach the target audience. Privacy experts were reached by advertising the study on the ICO's newsletter, while information systems and IT professionals were identified as members of the British Computer Society. As a result, a multiple-frame survey approach was adopted (Lohr 2009), which produced, as explained in the next section, acceptable rates of participation in the study.

5.6 Construct development: Formative Vs reflective measurement models

In measuring latent constructs, an important distinction to be made is the one between *formative* and *reflective* measurement models (MacKenzie, Podsakoff et al. 2011). Scale development procedures can be based either on the assumption that causality flows from the latent construct to the measures, in the sense that each measure is viewed as an imperfect reflection of the underlying latent construct – reflective models – or that indicators, rather than reflecting underlying latent constructs, can combine to form them – formative models (Bollen and Lennox 1991). Traditional examples of formative models in social sciences are "time spent with family and time spent with friends [which] are cause indicators of the latent variable of time in social interaction. Race and sex are cause indicators of exposure to discrimination" (Bollen 1989: p. 222).

Reflective measurement models are the most commonly used type of models in marketing and management studies (Diamantopoulos, Riefler et al. 2008). However, several scholars have begun advocating for the use of formative measurement models when appropriate (Bollen and Lennox 1991, Coltman, Devinney et al. 2008, Diamantopoulos, Riefler et al. 2008, MacKenzie, Podsakoff et al. 2011).

Examples of constructs measured with formative models are: 'organisational internet use' (Brock and Zhou 2005), and 'perceived effectiveness of institutional structures' (Pavlou and Gefen 2005) in the information technology literature. In marketing, constructs such 'CRM process implementation' (Reinartz, Krafft et al. 2004), 'sales forecasting effectiveness' (Winklhofer and Diamantopoulos 2002), 'service-oriented business strategy' (Homburg, Hoyer et al. 2002), 'marketing's influence' and 'market-related complexity' (Homburg, Workman et al. 1999) have been measured through formative scales. Formative measures have also been applied in management studies; examples include: 'firm reputation' (Helm 2005); environmental controls such as 'local government regulatory influence', 'quality of local business infrastructure', 'pressures of global competition' and 'pressures from technological change' (Venaik, Midgley et

al. 2005); ‘corporate identity’ and ‘corporate culture’ (Witt and Rode 2005); ‘drivers of a company's corporate reputation’ (Dowling 2004); ‘firm pressures’ and ‘corporate reputation’ (Venaik, Midgley et al. 2004); industry drivers such as ‘organisation structure’, ‘management process’ and ‘global strategy’ (Johansson and Yip 1994).

Reflective and formative measurement models differ (a) in the direction of construct-indicator causality, (b) in the path diagram, (c) in the way validity is assessed, (d) in the reasoning behind retaining or erasing indicators, (e) in the partitioning of the covariance matrix, and (f) in the identification method; though in both cases (g) linear composites of indicators can replace latent variables. A comparison and exhaustive overview of the differences between formative and reflective measurement models is reported in the Statistical Appendix.

Although the distinction between reflective and formative measurement models appears informative and compelling, no general rule prescribes whether a construct should be measured with a reflective or with a formative model. In fact “[c]onstructs are not inherently formative or reflective in nature” (MacKenzie, Podsakoff et al. 2011: p. 302). The same construct could even be measured in both ways depending on the way questions are formulated (Wilcox, Howell et al. 2008). A simple formative/reflective categorization may be overly simplistic and several criteria, including association, temporal precedence, and the elimination of rival causal explanations, should be employed (Edwards and Bagozzi 2000).

From a theoretical perspective, the fundamental question when deciding whether a construct should be measured in a formative or reflective way depends on whether individual items measuring the constructs share a common theme and can be considered interchangeable or not. In addition, if questions involve future actions a reflective model is more likely to be appropriate; the contrary is true if past actions are involved (Wilcox, Howell et al. 2008).

Survey items used in this study were composed of statements formulated in the present tense, which were referring to actions already initiated. Each statement was also measuring a particular sub-dimension of the construct; thus, statements were not interchangeable. Since assuming

reflective indicators to be formative indicators would produce a more serious problem than doing the opposite (MacKenzie, Podsakoff et al. 2005), the researcher has decided to treat all scales as formative ones. However, traditional validity and reliability tests used in the case of reflective measures have also been performed and reported in the Statistical Appendix in order to assess if a different decision would have produced different results. Thus, the rest of the discussion will focus on formative indicators. Four issues seem to be critical to the successful construction of formative indicators.

- *Content specification and indicator specification*: since under formative measurement the latent variable is determined by its indicators rather than vice versa, content specification is inextricably linked with indicator specification. Therefore, : items used as indicators must be sufficiently inclusive in order to capture fully the construct's domain of content to cover the entire scope of the latent variable (Diamantopoulos and Winklhofer 2001).
- *Indicator collinearity*: if a particular observed variable turns out to be almost a perfect linear combination of the other observed variables it can become a candidate for exclusion from the index because it is likely to contain redundant information (Bollen and Lennox 1991).
- *Validity*: nomological validity can be assessed by using a Multiple Indicators Multiple Causes (MIMIC) model, and/or structural linkage with another criterion variable (Coltman, Devinney et al. 2008).

While validity, defined as “the strength of the direct structural relation between a measure and a latent variable” (Bollen 1989: p. 222), can be applied to formative measures, the idea of reliability, which refers to the consistency of measurement, is more difficult to apply to cause indicators (Bollen 1989). Reliability in the internal consistency sense is not meaningful when a formative model is applied (Bagozzi 1994, Diamantopoulos and Winklhofer 2001).

Difficulties arising from identifying formative models with measurement error at the construct level often lead scholars to add two reflective indicators to the formative construct

(Diamantopoulos and Winklhofer 2001, MacKenzie, Podsakoff et al. 2005) and to rely on *Multiple Indicators and Multiple Causes* (MIMIC) methods, which represent a mixture of effect and causal indicators (Brown 2006). However, a problem known as 'interpretational confounding' can emerge as a result: the measurement model may change depending on what the central construct under study is predicting. Important differences in the way constructs are measured across studies might prevent accumulation of knowledge, because the version of the construct in each study might be incommensurable (Wilcox, Howell et al. 2008). Finally, both theoretical and empirical criteria are necessary to design and validate measurement models (Diamantopoulos 2005, Finn and Kayande 2005, Coltman, Devinney et al. 2008). The theoretical criteria which should be followed draw upon previous studies (Coltman, Devinney et al. 2008): (1) the nature of the construct; (2) the direction of causality between the indicators and the latent construct; (3) the characteristics of the indicators used to measure the construct.

Paralleling the three theoretical considerations above, are three empirical considerations that inform understanding of the measurement model: (4) indicator inter-correlation; (5) indicator relationships with construct antecedents and consequences; (6) measurement error and collinearity. Empirical considerations related to indicator inter-correlation are reported in the Statistical Appendix. Indicators' relationships with construct antecedents and consequences are explored in this chapter, in Part C. The following section will present the composite scores which have been used as a substitute for the latent constructs. Thus, the next section explains the way information gathered by each battery of questions has been summarised and transformed into continuous measures, each of them measuring a latent construct.

5.7 Construct operationalisation process: Construct dimensionality

5.7.1 Analytical Sophistication (SOPH)

In order to build analytical capabilities within an organisation, five elements, which are data, enterprise, leadership, targets, and analysts, must be present within the organisation (Davenport, Harris et al. 2010, Davenport 2014). These elements, which are part of the DELTA model presented in Chapter Two, section 2.5, characterised analytically sophisticated organisations. In order to measure the level of analytical sophistication an organisation has achieved, the researcher has developed a scale. Each item of the scale measures one dimension identified in the DELTA model. These dimensions are:

- **D for accessible, high-quality data**: data must be accurate, stored in compatible formats and easily accessible. They should be perceived to be a valuable asset within the organisation.
- **E for enterprise orientation**: the organisation need to have a flexible, centralized IT infrastructure enabling people to access and work with data.
- **L for analytical leadership**: in order to build an analytical culture within the organisation, employees should be encouraged by managers to rely on facts and data analysis.
- **T for strategic targets**: data analytics should build a distinctive, competitive capability to achieve targeted objectives within the organisation.
- **A for analysts**: the organisation must employ analysts able to mine data and generate useful business insights.

Recently, Davenport has proposed an extension of the DELTA model adapted to big data environments (Davenport 2014): it is called DELTTA model (data, enterprise, leadership, targets, technology, and analysts). The 'Big data Readiness Assessment Survey' (Brynjolfsson and McAfee 2013, Davenport 2014) was developed as an instrument to determine an organisation's readiness

for big data projects. This instrument, which can be applied to assess the "entire organisation or a business unit within it" (Davenport 2014: p. 206), was not available when the questionnaire for this study was developed, and, thus, survey items presented here do not match questions asked in the *Big data Readiness Assessment Survey*, whose content is reported in the appendix. Another difference to be noticed is that Davenport uses 5-point Likert scales in the *Big Data Readiness Assessment Survey*, while bipolar rating scales are used in this study.

The final instrument developed to assess an organisation's degree of analytical sophistication is reported in table 12. Questions Q21_1 and Q21_6 measure the 'Data' dimension; question Q21_2 measures the 'Enterprise' dimension; question Q21_3 measures the 'Leadership' dimension; question Q21_5 measures the 'Target' dimension; and, finally, question Q21_4 measures the 'Analysts' dimension.

Table 11. 'Analytical Sophistication' scale

CONCERNING THE ABILITY OF YOUR ORGANISATION TO ANALYSE AND MANAGE DATA EFFECTIVELY, could you please rate each of the following series of statements on a scale of 1 to 7, with opposing views at either end of the scale?

		1	**2**	**3**	**4**	**5**	**6**	**7**	
Q21_1	Within my organisation, data are accurate, stored in compatible formats and easily accessible.	○	○	○	○	○	○	○	Within my organisation, data are inaccurate, stored in incompatible formats and inaccessible.
Q21_2	My organisation has a flexible, centralized IT infrastructure to work with data.	○	○	○	○	○	○	○	My organisation lacks a flexible, centralized IT infrastructure to access and work with data.
Q21_3	Employees are encouraged to rely on data analytics.	○	○	○	○	○	○	○	Employees are not encouraged to rely on analytics-based knowledge.
Q21_4	We have analysts able to mine data and get useful insights.	○	○	○	○	○	○	○	We do not employ people with the necessary skills to analyse data.
Q21_5	Data analytics represents a distinctive, competitive capability of my organisation.	○	○	○	○	○	○	○	Data analytics does not build any competitive capability within my organisation.
Q21_6	Digital data represents a core asset, key to our business model.	○	○	○	○	○	○	○	Digital information does not add value to my organisation.

5.7.2 Dataveillance (DVEIL)

As discussed in Chapter Two, section 2.6, within this study "dataveillance specifically indicates the ability of reorienting, or nudging, individuals' future behaviour by means of four classes of actions: 'recorded observation'; 'identification and tracking'; 'analytical intervention'; and 'behavioural manipulation'" (Degli Esposti 2014: p. 210). Dataveillance involves the classification, compilation and analysis of vast amount of data of different kinds: from passenger information to financial transactions, people's attributes, preferences, online behaviour and so on. This kind of data is often, or is often linked to, personal information. As the data refer to identifiable or potentially identifiable people, the application of analytics to extract value out of this information is especially important in assessing the implications of big data analytics on organisational data protection practices. For this reason, dataveillance, here interpreted as organisational reliance on targeted analytics to foresee and influence human behaviour, represents a central element in understanding the articulation of big data components within organisations' information management practices.

The concept has three key dimensions, which are: monitoring, targeting, and nudging. Monitoring refers to the data gathering activity and the organisation of this information into digital records which can be analysed in search for pattern. Targeting refers to analytical procedures such as profiling, which can be equally used to identify criminal suspects and other risky groups (Levi and Wall 2004) or to identify valuable customers. Finally, nudging refers to the ability of influencing people's behaviour thanks to the knowledge created through the analysis of the data gathered. As it is difficult to assess whether an initiative has actually changed someone's behaviour, the question asked to measure this dimension refers only to an organisations' willingness to rely on analytics to influence someone's decisions. The questions, which have been used to measure the concept of dataveillance, are reported in table 13. The scale assesses the extent to which an organisation collects data to monitor individuals' activities, analyses personal data to foresee and influence people's behaviour, and relies on profiling to target valuable users or personalise offers.

Table 12. 'Dataveillance as Targeted analytics' scale

CONCERNING THE EXTENT TO WHICH YOUR ORGANISATION COLLECTS AND PROCESSES INDIVIDUALS' DATA, SUCH AS CUSTOMERS' DATA, could you please rate each of the following series of statements on a scale of 1 to 7, with opposing views at either end of the scale?

		1	2	3	4	5	6	7	
Q29_1	My organisation collects data to monitor individuals' activities.	❍	❍	❍	❍	❍	❍	❍	My organisation does not monitor people's activities through data collection.
Q29_2	We analyse personal data to foresee and influence people's behaviour.	❍	❍	❍	❍	❍	❍	❍	Anticipating and influencing individual behaviour is not an objective of data processing.
Q29_3	Profiling is used to target valuable users or personalise offers.	❍	❍	❍	❍	❍	❍	❍	Individuals' data are not analysed for profiling or segmentation purposes.

5.7.3 Data Pool Variety (DPOOL)

As introduced in Chapter Two, a diverse array of data is processed by the organisation; data such as: geographical location; unstructured data like voice, text or images; people's online behaviours, economic transactions, or individual attributes and attitudes. Data variety is considered a central characteristic of the big data phenomenon, with volume and velocity (Laney 2001). The accumulation of data related to any aspects of an individual's life has also been recognised in surveillance studies as a fundamental driver of change related to the proliferation of dataveillance practices (Lace 2005).

Questions Q33_1-6, included in table 14, measure the extent to which organisations are processing different categories of data. Question Q25, included in table 15, focuses on data volume.

Table 13. 'Data Pool Variety' scale

To what extent does your organisation analyse any of the following types of data? Please express your opinion on a scale from 0 = "Type of data not analysed" to 100 = "Type of data constantly analysed", by clicking on the graph.

		0	*10*	*20*	*30*	*40*	*50*	*60*	*70*	*80*	*90*	*100*
Q33_1	Data about people's online behaviours (e.g. click-streams; logs; search histories...)											
Q33_2	Data about geographical location (e.g. GPS or mobile telephone signals...)											
Q33_3	Unstructured data like voice, text or images (e.g. blogs; tweets; footages; videos...)											
Q33_4	Data about individuals' economic transactions (e.g. purchasing histories; credit cards operations...)											
Q33_5	Data about people's attitudes (e.g. survey opinions; "like" buttons...)											
Q33_6	Data about people's attributes (e.g. ethnicity; occupation; health conditions; sexual habits...)											

Table 14. 'Big data Volume' question

Q25. Do you know how much data your organisation is managing in your Big data environment today? (Only one answer allowed)

- ❍ I do not know
- ❍ 10 Terabytes or less
- ❍ 11 - 100 Terabytes
- ❍ 101 - 500 Terabytes
- ❍ 501 - 1 Pedabyte
- ❍ 2 Pedabytes or more
- ❍ None (not yet implemented a Big data environment)

5.7.4 Data Protection Regulatory Regime (REG)

Data protection regimes may strongly influence and even change organisational practices (Samiee 1984, Kane and Ricks 1988, Shaffer 1999). The way law is interpreted and enforced plays an important role in determining how norms influence organisational procedures and what effects

this causes on data privacy and security (Cockcroft 2003). A country's regulatory approach to the corporate management of information privacy is affected by its cultural values and by individuals' information privacy concerns, and most firms tend to take a reactive stand toward information privacy by waiting for strict regulation to be enforced before taking any action to reduce people's privacy concerns (Milberg, Smith et al. 2000).

Different cultures and nations develop and implement different privacy protection approaches (Milberg, Burke et al. 1995, Tang, Hu et al. 2008). The level of permissiveness of these approaches may determine its efficacy in setting minimum information privacy standards (Culnan 2000). The enforcement powers of regulators play an important persuasive role and raise business awareness of legal requirements (Ohlhausen 2014). The way the law is interpreted and enacted also influences the institutional environment and the marketplace in which an organisation operates (Campbell, Goldfarb et al. 2015).

As a result, question Q47_1 measures the degree of clarity and consistency of law and its interpretation. Q47_2 deals with data protection authorities' enforcement powers. Finally, question Q47_3 measures the dimension "preferences for regulation of information privacy" and was based on the following statement: "the best way to protect personal privacy would be through strong laws" (Milberg, Smith et al. 2000: p. 44). Questions Q47_1-3 are listed in table 16. Questions Q51_1-10 and Q52, which refer to the proposed General Data Protection Regulation and the impact of the provisions contained in the draft version published in 2014 (EC 2014), are reported in table 17.

Table 15. 'Data Protection Regulatory Regime' scale

Concerning data protection regulation in your country, could you please rate each of the following series of statements on a scale of 1 to 7, with opposing views at either end of the scale?

		1	2	3	4	5	6	7	
Q47_1	Data protection law is enforced in a consistent, reliable and predictable manner.	❍	❍	❍	❍	❍	❍	❍	Data protection law is enforced in an inconsistent, unreliable and unpredictable manner.
Q47_2	Data protection authorities have the power and the resources to impose serious sanctions if data are processed unlawfully.	❍	❍	❍	❍	❍	❍	❍	Data protection authorities do not have the power or resources to impose serious sanctions if data are processed unlawfully.
Q47_3	Tighter data protection regulations are necessary to ensure that all organisations meet minimum information security standards.	❍	❍	❍	❍	❍	❍	❍	Tighter data protection regulations are not needed to ensure organisations meet minimum information security standards.

Table 16. Assessment of the 'Provisions of the proposed General Data Protection Regulation'

According to your experience and taking into account the reality of your organisation, to what extent do you consider problematic implementing each of the following provisions envisioned by the proposed new European General Data Protection Regulation?

Q50_1	Serious data breaches must be notified to both the Data Protection Agency and data subjects. Supervisory authorities will maintain a public register of the types of breach notified. Notification must be given without undue delay.	❍ Highly problematic ❍ Somewhat problematic ❍ Not very problematic ❍ I do not know
Q50_2	Consent must be given by a data subject in a clear statement or via an affirmative action (i.e. ticking a consent box when visiting a website) in cases when explicit consent would be required.	❍ Highly problematic ❍ Somewhat problematic ❍ Not very problematic ❍ I do not know
Q50_3	A data protection officer (DPO) must be appointed by public authorities and businesses if data of more than 5000 data subjects is processed in any consecutive 12-month period. A DPO will also have to be appointed if (i) special categories of data, (ii) location data, (iii) data relating to children, or (iv) employee data in large scale filing systems are processed.	❍ Highly problematic ❍ Somewhat problematic ❍ Not very problematic ❍ I do not know
Q50_4	Data Protection Impact Assessment (PIA) must be performed annually. Companies are also encouraged to adopt Privacy by Design principles (PbD) and to certify their data processing by a supervisory authority, possibly in cooperation with accredited third party auditors.	❍ Highly problematic ❍ Somewhat problematic ❍ Not very problematic ❍ I do not know
Q50_6	Data subjects will have the right to data portability, which is a right to require a portable copy of a data subject's personal data so that they may transfer it to another data controller.	❍ Highly problematic ❍ Somewhat problematic ❍ Not very problematic ❍ I do not know

Q50_8	Data subjects will have the right to erasure. This will allow individuals to have all personal data that business holds on them deleted or restricted. This will include all photos and any public links to, or copies of, personal data that can be found on the Internet for example in social networks or via search engines.	❍ Highly problematic ❍ Somewhat problematic ❍ Not very problematic ❍ I do not know
Q50_10	The regulation will apply to organisations outside the EU whenever they process personal data of individuals in the EU. Data transfer outside the EU will be possible through Binding Corporate Rules (BCR) or in case of authorisation given by data protection authorities. Authorisations will be valid only for two years.	❍ Highly problematic ❍ Somewhat problematic ❍ Not very problematic ❍ I do not know
Q50_9	Other (please specify):	❍ Highly problematic ❍ Somewhat problematic ❍ Not very problematic ❍ I do not know

Table 17. 'Proposed General Data Protection Regulation': Organisational readiness

Q52. The draft Regulation still needs to be approved by the member states and ratified by the European Parliament before it can be adopted. It is expected that this process will take approximately two/three years. Most privacy lawyers expect there to be major changes to data protection legislation, with many of the provisions of the draft GDPR being implemented.

Given its likely impact, has your organisation started planning for the new Regulation?

- ❍ Yes
- ❍ No
- ❍ I do not know
- ❍ Other (Please specify): _____

5.7.5 Compliance with Data Controllers' Obligations (DPP)

As explained in section 3.3 and in section 4.4, according to European data protection law (EC/46 1995), organisations operating in Europe have to comply with basic data controllers' obligations. According to these obligations, organisations are expected to: keep data complete, accurate and up-to-date; collect the minimum amount of data necessary to fulfil a specific objective; delete data once the objective for which they have been collected was achieved; shares individuals' data only with authorised third parties; establish strong security measures to protect data from unauthorised use; sanction those who use or handle personal data inappropriately; and establish procedures to compensate individuals in case personal data were compromised.

Questions Q31_1-4 and Q41_1-3 assessed the extent to which, from the perspective of the respondent, a specific organisation was compliant with each obligation (see table 19). Because of the large number of obligations, questions were divided in two batteries.

Table 18 'Compliance with Data Protection Principles' scale

Please indicate which statement better reflects your organisation's approach to the management of individuals' data, on a scale of 1 to 7, with opposing views at either end of the scale.

		1	2	3	4	5	6	7	
Q31_1	We keep data complete, accurate and up-to-date.	❍	❍	❍	❍	❍	❍	❍	We deal with partial, inaccurate and outdated data.
Q31_2	We try to collect the minimum amount of data necessary to fulfil a specific objective.	❍	❍	❍	❍	❍	❍	❍	We collect as much data as we can to fulfil new objectives.
Q31_3	We delete data once the objective for which they have been collected is achieved.	❍	❍	❍	❍	❍	❍	❍	We retain data indefinitely in case of future use.
Q31_4	We only share individuals' data with authorised third parties.	❍	❍	❍	❍	❍	❍	❍	We share individuals' data with any third party.

Please indicate which statement better reflects your organisation's approach to secure personal data, on a scale of 1 to 7, with opposing views at either end of the scale.

		1	2	3	4	5	6	7	
Q41_1	We sanction those who use or handle personal data inappropriately.	❍	❍	❍	❍	❍	❍	❍	No internal policy foresees any sanctions for who uses personal data inappropriately.
Q41_2	Strong security measures protect data from unauthorised use.	❍	❍	❍	❍	❍	❍	❍	We do not have specific security measures to protect data from unauthorised use.
Q41_3	We have procedures in place to compensate individuals in case data were lost, manipulated or stolen.	❍	❍	❍	❍	❍	❍	❍	Individuals will get no compensation in case anything goes wrong with their data.

5.7.6 Respect of Data Subjects' Rights (DSR)

As explained in section 3.3 and in section 4.4, according to data protection law (EC/46 1995), organisations operating in Europe have to take the appropriate measures to ensure the respect of data subjects' rights. According to these principles, individuals must be fully informed about all aspects related to the processing of their data; when appropriate, Individuals should be asked to

give their explicit consent to the processing of their data; they must also have procedures in place to let individuals rectify inaccurate data; organisations must have procedures in place to satisfy individuals' requests to end the processing of their data.

Questions Q37_1-4 have been used to assess the extent to which organisations, according to respondents' perceptions, are taking all the necessary means to ensure data subjects' rights are respected (see table 20).

Table 19. 'Respect of Data Subjects' Right' scale

Please indicate which statement better reflects the way your organisation handles personal data, on a scale of 1 to 7, with opposing views at either end of the scale.

		1	2	3	4	5	6	7	
Q37_1	We always obtain explicit consent from individuals before processing their data.	❍	❍	❍	❍	❍	❍	❍	We do not rely on consent but on alternative means for processing personal data (e.g. the processing is necessary in relation to a contract).
Q37_2	Individuals are fully informed about all aspects related to the processing of their data.	❍	❍	❍	❍	❍	❍	❍	We do not have to inform individuals about all aspects related to the processing of their data.
Q37_3	We can easily satisfy individuals' requests to end the processing of their data.	❍	❍	❍	❍	❍	❍	❍	We cannot satisfy individuals' requests to stop the processing of their own data.
Q37_4	We have procedures in place to let the individuals rectify inaccurate data.	❍	❍	❍	❍	❍	❍	❍	There are no means for rectifying inaccurate data on the basis of an individual request.

5.7.7 Organisational Privacy Culture (PRV)

As discussed in Chapter Four, section 4.5.3, the development of a strong information protection culture is extremely important to ensure personal data are not misused or manipulated (Da Veiga and Martins 2015). The respect for consumers' privacy is also an ethical imperative for marketers who need to make an active commitment to ethical behaviour if they want to safeguard consumers' trust in online commerce (Jones 1991, Foxman and Kilcoyne 1993, Bernard and Makienko 2011). When the respect for privacy is translated into an organisational value, it can also be perceived as a positive quality attribute and adding value to the organisation (Storey, Kane et al. 2009). High ethical and privacy standards can also contribute to create a safer information

management environment where good data handling procedures are more likely to be followed (Culnan and Williams 2009). As a result, the emergence of a strong privacy culture within the organisation should lead the organisation to: transform privacy into a distinctive organisational feature; establish the respect for information privacy as a core organisational value; and invest considerable human and economic resources in securing information. Questions Q35_1-3 reflect and assess—from the respondent's perspective—the extent to which these ideas are present within the respondent's organisation. In the first question "privacy represents a distinctive feature of my brand/organisation" the researcher uses the term "brand" and "organisation" interchangeably to help respondents, working in the for-profit and nonprofit sector, understand the question.

Table 20. 'Organisational Privacy Culture' scale

Concerning your organisation's approach to individuals' privacy, such as customers' privacy, and information security, could you please rate each of the following series of statements on a scale of 1 to 7, with opposing views at either end of the scale?

		1	**2**	**3**	**4**	**5**	**6**	**7**	
Q35_1	Privacy represents a distinctive feature of my brand/organisation.	❍	❍	❍	❍	❍	❍	❍	Privacy is not one of my brand/organisation's distinctive features.
Q35_2	Remarkable human and financial resources are devoted to secure information.	❍	❍	❍	❍	❍	❍	❍	Almost no human or financial resources are dedicated to information security.
Q35_3	Privacy is a core value, central to our organisational culture.	❍	❍	❍	❍	❍	❍	❍	Privacy does not represent an essential part of the organisational culture.

Table 21. Questions on 'Frequency of Data Breaches'

Q44. On the base of your knowledge, how often do organisations in your sector experience serious breaches of personal data? *(Only one answer allowed)*

- ❍ Incidents may occur on a daily basis
- ❍ Incidents may occur on a weekly basis
- ❍ Incidents may occur on a monthly basis
- ❍ Incidents may occur on a yearly basis
- ❍ I have never heard of any incident in my sector
- ❍ I do not know

Table 22. Questions exploring common causes of data breaches

In your opinion, what most commonly causes data breaches? (More than one answer allowed)	
Q45_1	Unintended disclosure (e.g. sensitive information posted publicly on a website or sent to the wrong party via email).
Q45_2	Lost, discarded or stolen stationary electronic device (e.g. desktop computers, servers...).
Q45_3	Hacking, malwares or spywares.
Q45_4	Lost, discarded or stolen non-electronic records (e.g. paper documents).
Q45_5	Payment Card Fraud (e.g. skimming devices at point-of-service terminals).
Q45_6	Insiders (someone with legitimate access—such as an employee or contractor—who intentionally breaches information).
Q45_7	Lost, discarded or stolen portable device (e.g. laptop, PDA, smart-phones, USB, CDs...).
Q45_8	I do not know.
Q45_9	Other (please specify):

Table 23. Questions on 'Privacy and Security Safeguards Adopted'

Which privacy or security safeguards has your organisation already adopted? (More than one answer allowed)	
Q53_1	A Chief Privacy/Data Protection Officer is in charge of supervising all privacy-related issues.
Q53_2	The function of dealing with privacy-related matters is pursued by a designated department inside my organisation, for example the compliance office or the IT department, etc.
Q53_3	Data policies that describe the rules controlling the integrity, security, quality, and use of data during its life-cycle and state change have been adopted.
Q53_4	Specific policies for classifying information according to their sensitivity (e.g. secret; confidential; for internal use; etc.) are in place.
Q53_5	Consent obtained through opt-in acceptance of data processing terms and conditions.
Q53_6	Consent obtained through opt-out acceptance of data processing terms and conditions.
Q53_7	Employees are constantly trained to comply with privacy procedures.
Q53_8	Workforce members are sanctioned if they do not comply with privacy procedures.
Q53_9	Privacy Enhancing Technologies (PETs) are in use.
Q53_10	Privacy-by-design (PbD) criteria are adopted in product development.
Q53_11	Privacy Impact Assessments (PIAs) are undertaken.
Q53_12	Counsel of a legal firm specialized in information privacy.
Q53_13	Binding Corporate Rules (BCRs) to manage international data transfer.
Q53_14	Periodical external auditors' assessment of internal security standards.
Q53_15	Immediate notification to individuals if their data are breached disclosed or manipulated.
Q53_16	Certified code of practice for information security management (e.g. ISO/IEC 27002:2005).
Q53_17	Data breach insurance policy.
Q53_18	Full-disk encryption of physical devices like laptops or PCs.
Q53_19	Encrypted transmission of data.
Q53_20	Network and application penetration and vulnerability testing (e.g. friendly hacking).
Q53_21	Other (please specify): _____
Q53_22	I do not know

5.7.8 Use of Analytics across Business Functions (FUNCT)

As already discussed in Chapter Two, section 2.4, big data analytics is present in every industry and it can be applied to solve problems virtually within any business function (Davenport 2014). "Marketing, sales, supply chain, manufacturing, human resources, strategy, finance, information technology" are some key areas where big data analytics can have a big impact (Davenport 2014: pp. 50-56). For this reason, questions Q26_1-7 assess whether analytics is used by organisations to:

1) foster marketing;
2) improve security;
3) gain efficiency;
4) manage human resources;
5) reduce financial risks;
6) take better informed strategic decisions.

The measurement scale used goes from "0%" to "100% used for this purpose", which generated an interval variable (see table 26).

Table 24. 'Functional Use of Analytics' scale

To what extent does your organisation use **data analytics** to pursue your organisational goals in any of the following areas? Please express your opinion on a scale from 0 = "Not used" to 100 = "Definitely applied for this purpose".

		0	*10*	*20*	*30*	*40*	*50*	*60*	*70*	*80*	*90*	*100*
Q26_1	To foster marketing											
Q26_2	To improve security											
Q26_3	To gain efficiency											
Q26_4	To better manage human resources											
Q26_5	To reduce financial risks											
Q26_6	To take better informed strategic decisions											
Q26_7	To offer public policy services											

5.8 Exploratory analysis: Antecedents of information security investment decisions

High profile data breaches can increase information risk perceptions (Volpentesta, Ammirato et al. 2011) and have an effect on the budget for security expenditures within organisations (Bikard 2011). Security threat awareness and immediacy may contribute to motivate employees to comply with security policies (Siponen, Pahnila et al. 2010). Nonetheless, famous cases, such as the Sony BMG Rootkit incident in 2005, demonstrate a failure to adequately value security and privacy as part of a firm's strategy (Mulligan and Perzanowski 2007). To address information security threats, a risk management approach, increasingly demanded both by firms and regulators (Bamberger 2010), is often adopted to determine the optimal level of investments in customer information security (Lee, Kauffman et al. 2011).

Information security vulnerabilities and data breaches create disruptions inside and outside the organisation. They threaten consumers' trust in online commerce and may produce serious economic losses on the stock market (Spanos and Angelis 2016). As a result, large companies involved in e-commerce pay special attention to their information security performance and tend to disclose this information (Li 2015). Incidents can also trigger regulators' enforcement actions and impose serious sanctions to organisations which have failed to secure their information. However, besides reducing risks, investing in information security brings other positive effects. It leads to positive market returns (Chai, Kim et al. 2011), contributes to the adoption of best practices and quality procedures (Mesquida and Mas 2015), and generates a better understanding of malicious behaviour (Crossler, Johnston et al. 2013).

To assess the motivations behind information security investment decisions, questions Q43_1-8 present different options to the respondent (see table 22). Organisations may invest in information security to:

1) react to previous security problems;

2) manage the risk of high litigation costs;

3) avoid costly enforcement action by regulators;

4) manage the risk of economic loss;

5) manage reputational risks;

6) reflect best information security practices;

7) improve service or product quality.

Because of the importance of data breaches in influencing market reactions, as presented in sections 3.8 and 4.7, questions Q44 and Q45_1-9 explore the frequency of occurrence of data breaches and the most common factors causing them (see table 23 and table 24). In addition, questions Q53_1-22 present specific privacy and security measures which could have been adopted by the organisation (see table 25).

Table 25. Questions on 'Information Security Investment Decisions'

In general, what motivates investments in information security (InfoSec) inside your organisation? Please express, on a scale from 0 = "It is not at all a relevant reason to invest in InfoSec" to 100 = "It is a very relevant reason to invest in InfoSec", how relevant each factor is for your organisation

		0	*10*	*20*	*30*	*40*	*50*	*60*	*70*	*80*	*90*	*100*
Q43_1	To manage the risk of high litigation costs											
Q43_2	To reflect high industry information security standards											
Q43_3	To manage the risk of economic loss											
Q43_4	To manage reputational risks											
Q43_5	To improve service/product quality											
Q43_6	To avoid costly enforcement action by regulators											
Q43_7	To react to previous security problems											
Q43_8	I do not know											

5.9 Organisational characteristics

Some industries are more well suited to Big data than others (Davenport 2014). Internet companies and firms operating in business-to-consumers sectors tend to be 'overachievers' with big data. Other companies, such as telecom firms, media and entertainment firms, retailers, traditional banks and electric utilities can be considered 'underachievers', as they have data but do not know how to use them effectively (Davenport 2014: pp. 42-44). Finally, companies operating in business-to-business sectors, industrial product firms or firms which have too many intermediaries and no data about their final clients, are considered to be "data disadvantaged organisations", as they do not have enough or well-structured data (Davenport 2014: p. 43).

The same happens with information security investment decisions and compliance with data protection laws. Small and medium-size companies often do not have the resources to address regulatory requirements (Goucher 2011, Kurpjuhn 2015). Regulation also varies between countries and, within countries, it may vary between industries. For profit organisations are subject to different rules than nonprofit firms, and firms selling on the business-to-business market are less exposed to changing public privacy concerns than firms selling directly to consumers.

For all these reasons, it is important to take into consideration basic organisational characteristics in order to correctly interpret results. Questions included in the questionnaire, which assess different organisational features, are reported in table 27.

Table 26. Questions on organisational characteristics

Q4	How many staff does your organisation, or major contractor, employ?
Q7	Is the office where you are based located in the United Kingdom?
Q8	What is the location of the office where you are based?
Q9	Could you please specify where your office is based in the United States?
Q10	Is the organisation where you work a for profit or a nonprofit institution?
Q11	Which of the following best describes the sector in which your company operates?
Q12	Which of the following best describes the sector in which your organisation operates?
Q13	What is your organisation's annual turnover (please refer to 2012 revenue)?
Q14	What is your organisation's annual budget (please refer to 2012 revenue)?
Q15	Compared with 24 months ago, has your organisation's turnover increased, decreased or stayed roughly the same?

Q16	Compared with 24 months ago, has your organisation's budget increased, decreased or stayed roughly the same?
Q17	In the next 12 months do you expect your organisation' turnover to increase, decrease, or stay roughly the same?
Q18	In the next 12 months do you expect your organisation' budget to increase, decrease, or stay roughly the same?
Q19	Taking into account all sources of income in the last financial year, did your firm generate a profit or a surplus?
Q39	Who is your company's typical customer?
Q40	Who is your organisation's typical end user?
	DATA INTENSIVE BUSINESS MODELS
	Does your organisation do any of the following?
	FOR PROFIT
Q27_1	My organisation promotes or sells its products or services on Internet.
Q27_2	My organisation uses monitoring devices to track customers or other people (e.g. web cookies, RFID, smart CCTV).
Q27_3	My organisation generates income by storing data for other organisations.
Q27_4	My organisation generates income by selling data.
Q27_5	My organisation generates income by analysing data.
Q27_6	My organisation is an ISP, hosting or cloud provider.
Q27_7	My organisation is in the online advertising business.
Q27_8	None of the above.
	NONPROFIT
Q28_1	My organisation promotes its services through a website.
Q28_2	My organisation organises fund raising campaigns on Internet.
Q28_3	My organisation uses monitoring devices to track users or other people (e.g. web cookies, RFID, smart CCTV).
Q28_4	None of the above.

5.10 Respondents' characteristics

Questions about respondents' characteristics were also included in the questionnaire (see table 28). These questions are meant to assess how knowledgeable respondents were about certain aspects of the organisation, such as information management practices and analytics.

Table.27. Questions on respondents' characteristics

Q5	How long have you been unemployed?
Q23_1	Do you feel you have the necessary statistical and computational skills to analyse data? ("I would have no idea where to start ⇔ I am 100% a data analyst")
Q24_1	How knowledgeable are you about your organisation's Information Systems Management practices? ("I have no idea ⇔ I am very knowledgeable")
Q49_1	Do you feel you have the necessary knowledge and legal skills to understand data protection laws? ("I have no idea ⇔ I am a data protection expert")
Q55	What is your job title?
Q56	Overall, how many years of working experience do you have?

Q57	What is your educational background?
Q57_TEXT	What is your educational background? - TEXT
Q59	We would like to discuss some issues further with you and answer your questions. WOULD YOU BE INTERESTED IN TAKING PART IN THE FOLLOW-UP OF THE STUDY? If you say 'yes', our researchers will contact you for arranging an interview.
Q61	Any final comment?

5.11 Questionnaire development

As presented in chapters two, three and four, an extensive literature review was carried out in order to map and understand each construct's domain and dimensionality. In this chapter, each construct was operationalised into questions and organised within batteries of questions mapping the same construct (MacKenzie, Podsakoff et al. 2011). Questions which were part of the same scale appeared jointly in the electronic questionnaire, which was developed and administered through a web-based service called *Qualtrics*.

Questionnaire items were tested during two pilot studies. As part of pilot number one, a paper-based version of the questionnaire was sent to two security experts, one marketing practitioner and one information management practitioner, two scholars and several representatives of the organisations supporting the media distribution strategy in order to assess the clarity of the wording of survey items, their exhaustiveness, and the order of appearance of questions. The panel offered feedback by commenting on each question. As part of pilot two, members of the UK *National Association of Data Protection Officers* (NADPO) revised a preliminary version of the electronic questionnaires. This version of the instrument contained specific text-boxes to let panel members give their feedback. Comments and characteristics of experts are discussed in this chapter, in section 5.11.2, and also reported in the Methodological Appendix.

A summary of the various stages of the questionnaire development process is reported in table 28. The constructs' definition phase was carried out in Chapter Four. Details on instrument's and indicators' development are reported in this chapter. Considerations on sample' composition are reported in Chapter Six.

Table 28. Instrument development and validation process

	Instrument development stages	Dissertation's chapters
	STAGE A: Constructs' definition	**Chapter 2, 3 and 4**
1	Rationale of the study and elaboration of the research question.	
2	Specification of the definition, domain and dimensionality of constructs through an extensive literature review of previous studies in the area.	
3	Construction of a theoretical model and relationships between constructs	
	STAGE B: Instrument's development	**Chapter 5 and Methodological Appendix**
4	Production of a sample of items meant to measure constructs' dimensions.	
5	Creation of draft instrument through repeated testing iterations. Assessment of overall instrument quality: Pilot 1 – expert judges.	
6	Assessment of content validity: Pilot 2 – data protection experts.	
	STAGE C: Indicators' development	**Chapter 5 and Methodological Appendix**
7	Consideration related to scales' characteristics, reliability, and constructions of indicators measuring each construct.	
8	Consideration related to scales' validity (construct validity; convergent and discriminant validity; adequacy of model's fit; concurrent validity; nomological validity).	
	STAGE D: External validity assessment	**Chapter 5, 6 and 7**
9	Identification of the target audience and elaboration of survey distribution strategy.	
10	Generalizability of results.	

Source: Author's elaboration of (Smith, Milberg et al. 1996).

5.11.1 Measurement scales

In order to measure the constructs presented in previous chapters, various semantic differential scales were developed, following previous information system studies (Verhagen, van Den Hooff et al. 2015). Also in line with previous marketing studies (Meadows and Dibb 2012), a one-to-seven scale, with opposite statements at each end of the scale, was used to measure the internal dimensions of each construct. One-to-seven Likert scales (1 = strongly disagree, 7 = strongly agree) were also used in previous privacy studies (Milberg, Smith et al. 2000).

The one-to-seven scale was chosen since it was demonstrated that bipolar rating scales with seven points yield measurement accuracy superior to that of three-, five-, and nine-point scales (Malhotra, Krosnick et al. 2009). Numbers, rather than adverbs, were preferred in order to both indicate the distance on the semantic scale and to avoid the heterogeneous interpretation of the meaning of each word and polarisation at the extreme of the scale (Dolch 1980). Entire phrases,

rather than simple adjectives, were also preferred in constructing survey items (Dickson and Albaum 1977).

All constructs, with definitions, internal dimensions and survey items, are presented within this section. All multidimensional constructs were measured by seven-point bipolar measurement scales (see table 29). These constructs are: Analytical Sophistication; Dataveillance; Data Protection Regulatory Regime; Compliance with Data Protection Principles; Respect for Data Subjects' Rights; and Organisational Privacy Culture.

The scale used to measure unidimensional constructs was a scale based on percentages, which goes from zero to one-hundred percent, where 100% indicates full adoption or use, and 0% indicates no adoption or use. Thus, scales based on percentages were used to measure the following constructs, which are: Data Pool Variety; Use of Analytics across Business Functions; and Rationale behind Investing in Information Security. The 0%-100% scale generated interval data, which are "numeric data where the intervals between values have meaning" because intervals have the same size (Linebach, Tesch et al. 2014: p. 396).

Table 29. List of constructs with measurement scales

Big data Constructs		***Data Protection Constructs***	
Multidimensional constructs (1-7 bipolar scale)	*Unidimensional constructs (0%-100% scale)*	*Multidimensional constructs (1-7 bipolar scale)*	*Unidimensional constructs (0%-100% scale)*
1. Analytical Sophistication	2. Data Pool Variety	5. Data Protection Regulatory Regime	6. Rationale behind Investing in Information Security
3. Dataveillance as Targeted analytics	4. Use of Analytics across Business Functions	7. Compliance with Data Controllers' Obligations	
		8. Respect for Data Subjects' Rights	
		9. Organisational Privacy Culture	

An advantage of relying on a one-to-seven scale is that it could be treated as an interval variable, though it generates an ordinal variable. As a variable's measurement scale determines which statistical methods are appropriate (Agresti 2013), treating these variables as interval let the

researcher use statistical techniques developed for performing quantitative data analysis. In the next section association tests will be used to explore the relationship between variables.

In summary, the survey instrument mostly contained closed-answer questions meant to produce ordinal, interval and ratio variables. Dichotomous variables and categorical variables, measuring dimensions such as firm industry, location or respondent's education were also included in the survey instrument.

5.11.2 Questionnaire pre-testing

Three tests were run to ensure linguistic clarity, psychological bipolarity, and unidimensionality of the concepts expressed in each question (Verhagen, van Den Hooff et al. 2015). First, a pretest for linguistic contrast with native speakers was initially performed. Then, tests for psychological bipolarity with experts, judging the linguistic alignment of each bipolar scale in relation to the concept under study, were undertaken during the second pilot. During this pilot, two information security experts, one marketing expert and one financial accounting expert were asked to revise the questionnaire. On the base of their feedback, various questions were amended. Questions asking for an organisation's annual revenue and workforce, for instance, had too many categories, and had to be reduced. A brief explanation of the scope of the study and of key terms was added at the beginning of the survey to ensure everyone understood the objective and terminology adopted. A specific question to identify internet and telecommunications providers was also introduced. Questions that were felt to be too general were identified and their formulation revised. Some overlapping categories were identified in the specific question which measure work experience and it was erased.

Finally, a group of expert judges revised the questionnaire to ensure face validity, which refers to logical or conceptual validity (i.e. prima facie evidence), as well as content validity, which refers to the ability of measurement items to accurately represent the meaning of the concept being measured. Between the 3rd and the 27th of June 2013, eleven members of *The National Association of Data Protection Officers* (NAPDO) commented on the questionnaire and helped

improve questions' clarity and quality: eight of those who contributed to test the instrument said that they worked for non-profit sector organisations; while only three of them worked for for-profit organisations. NAPDO members had between 5 and 20 years of working experience, and most of them were *Heads of Information Governance*, dealing with privacy, information security and data protection issues.

Thanks to their feedback, the researcher modified some questions to better reflect the reality of nonprofit organisations. Wording was further refined and sentences were made shorter to decrease reading time. As judges were also filling in the questionnaire while writing their comments on boxes inserted in the electronic questionnaire for the occasion, the researcher was also able to check if there was sufficient variability in the distribution of responses. All comments given and changes made are reported in the Methodological Appendix.

5.12 Construct operationalisation process: Unweighted composite scores

Regardless of whether we consider that a person's score on a measure of a latent construct is a function of his/her true position on the latent construct, plus error, or that we conceive constructs' meaning as emanating from the indicators to the construct in a definitional sense, "linear composites of indicators can replace latent variables" (Bollen and Lennox 1991: p. 305). Furthermore, similarities between linear composites of effect indicators and causal indicators are intriguing: in the presence of measurement error, or equation disturbance, the linear composite of both cases has less than perfect correlation with the latent variable; in both cases also this correlation can be altered by weighting the indicators (Bollen and Lennox 1991).

Computing the sum of equally weighted items to form a composite scale to measure the latent variable is a common practice. However, "[t]he appropriateness of the linear composite, like other conventions, depends on whether the latent variable is measured with effect or causal indicators" (Bollen and Lennox 1991: p. 309). When indicators of a construct present unique aspects of the construct, the construct can be viewed as a sum, or a composite, of the individual

indicators (Bagozzi 1994, Homburg, Hoyer et al. 2002). Because each of these items measures a particular dimension of the underlying construct, they all contribute to the total value of the corresponding construct. As a result some scholars consider that it would be more appropriate to talk about formative indicators when a latent variable is defined as a linear sum of a set of measurements (Diamantopoulos and Winklhofer 2001). As a formative specification implies the following relationship between the observed variables and the latent variable (Diamantopoulos and Winklhofer 2001), the constructs presented in the previous chapter have been treated and measured as linear composite variables of the following type (Mulaik 2009):

$$Y = a_1X_1 + \ldots + A_nX_n$$

where

Y is a linear composite variable

X_i 's are component random variables

A_i 's are numerical constants that serve as "weights" indicating by how much the scores on the components are multiplied, respectively, before entering the composite.

By assuming that – without loss of generality – each variable has the same weight (a = 1), we obtain a simple, unweighted composite, which are composite variables that do not weight the component variables differently (Mulaik 2009). Unweighted composite variables are easier to interpret and to be replicated in future studies. In contrast, factor scores always vary from study to study preventing measurements being exactly replicated. In addition, in the case of the construct Analytical Sophistication scholars recommend "averaging the scores within each DELTTA factor to create an overall score" (Davenport 2014: p. 203).

By following this procedure, nine indexes were constructed to measure the constructs. Descriptive statistics of each of them are reported in table 30, while indexes' probability distributions are displayed in the Statistical Appendix. Standardised indexes, with mean equal to zero and variance equal to one, were also produced to avoid problems with differences in

measurement scales. The direction of the variables was also adjusted to ensure all variables moved from low to high values.

Table 30. Descriptive Statistics of the original and standardised indicators

	Descriptive Statistics											
	Original indicators				Standardised indicators				Skewness		Kurtosis	
	Min	Max	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Stat.	Std. Error	Stat.	Std. Error
1. SOPH	-7	0	-3,32	1,32	-2,65	2,51	0	1	-,06	,17	-,48	,35
2. DVEIL	-7	0	-3,77	2,32	-1,39	1,63	0	1	,12	,17	-1,22	,35
3. DPP	-6	0	-2,57	1,40	-2,25	1,84	0	1	,07	,17	-,45	,35
4. PRV	-7	0	-2,43	1,72	-2,66	1,41	0	1	-,71	,17	-,12	,35
5. DSR	-7	0	-2,64	1,69	-2,59	1,56	0	1	-,34	,17	-,32	,35
6. REG	-7	0	-2,48	1,82	-2,48	1,37	0	1	-,72	,17	-,04	,35
7. DPOOL	0	100	26,48	21,96	-1,21	3,35	0	1	1,02	,17	,84	,35
8. FUNCT	0	100	46,10	24,44	-1,89	2,21	0	1	-,12	,17	-,51	,35
9. ISEC	0	100	59,91	21,07	-2,84	1,90	0	1	-,50	,17	,27	,35
Valid N	195											

Key: Analytical Sophistication (SOPH); Dataveillance as Targeted Analytics (DVEIL); Compliance with Data Protection Principles (DPP); Organisational Privacy Culture (PRV); Respect of Data Subjects' Rights (DSR); Data Protection Regulatory Regime (REG); Data Pool Variety (DPOOL); Functional Use of Analytics (FUNCT); Information Security Investment Decisions (ISEC).

As explained in the following section, these composite scores have been used to investigate the object of inquiry and give an answer to the research questions. The technique used, called Structural Equation Modelling, can be considered a fundamental part of the research process, rather than a mere statistical tool (Bollen 1989). The construction of the model started with its conceptualization in Chapter Four. Constructs were operationalised and the instrument was developed in Chapter Five.

5.13 Data analysis

5.13.1 Preliminary data check methods

Information on survey drop-out rates will be provided as in the case of electronic surveys quite often respondents stop answering questions because of fatigue or time constraints. To ensure that the source from which participants received the invitation to take part in the study has not

produced some sort of major distortion, or selection bias, the nonparametric Kruskal-Wallis H Test (Kruskal and Wallis 1952, Kruskal and Wallis 1953) will also be used to test whether the distribution of survey responses varies depending on the media channel publishing information about the study. The H test is considered the nonparametric alternative to the one-way ANOVA, and an extension of the Mann-Whitney U test to allow the comparison of more than two independent groups (Wasserman 2006). It is appropriate to use this test when data are ordinal and observations are independent. The null hypothesis is that the mean ranks of the *k* groups will not substantially differ.

The H test can be performed to test whether the distribution of some key variables were the same, or varied, across survey distribution channels (Linebach, Tesch et al. 2014). In addition, To ensure that no participant belonged to more than one group (Siegel 1956), records will be individually checked. Although the survey was anonymous, meta-data such as IP addresses and information about respondents' location latitude and longitude were collected by Qualtrics to ensure that the same person was not completing the survey more than one time.

5.13.2 Test of hypotheses with path analysis

Path analysis is a type of multiple-regression with relies on Structural Equation Modelling (SEM) technique but without the presence of measurement errors (Grapentine 2000). SEM is widely used in business studies, especially in marketing studies (Babin, Hair et al. 2008), but also in information systems research (Chin 1998, Urbach and Ahlemann 2010), management studies (Davcik 2014), strategic management (Shook, Ketchen et al. 2004), and management accounting (Smith and Langfield-Smith 2004). In general, SEM defines a set of data analysis tools that allows the researcher to test whether a set of independent variables and a set of dependent variables – either continuous or discrete – are related according to *a priori* specified hypotheses (Skrondal and Rabe-Hesketh 2004). SEM provides unbiased estimation when it is reasonable to consider that some external factors could have influenced the loss of information and the generation of missing data (Muthén, Kaplan et al. 1987).

No simple rule determines optimal sample size when using SEM: the appropriate size varies depending on the researcher's objectives (Fabrigar, Porter et al. 2010), and ad-hoc algorithms could be necessary to compute the appropriate sample size while taking into consideration several features of the model (Christopher Westland 2010). Depending on the number of variables in the model, SEM can be applied with a fairly small sample – 100 cases for example – still ensuring reliable results (Iacobucci 2010). However, a small sample size (for example n = 200) may not meet the recommended ratio of sample size to number of free parameters (5 to 1) needed to obtain trustworthy parameter estimates in structural equation modelling (Baumgartner and Homburg 1996). A solution is to compute unweighted composite scores to measure each construct and rely on path analysis. The main advantage is that path analysis requires the estimation of an inferior number of parameters with respect to SEM (Dow, Wong et al. 2008). However, the fact of relying on composite scores, rather than on the original observed variables, has the drawback of causing a partial loss of information.

In selecting the estimation method to compute model parameters considerations related to the normality of the multivariate empirical distribution of the data need to be made. If the multivariate distribution is leptokurtotic, with positive excess kurtosis (7.02 +/- 3.88), relying on GLS estimation is not appropriate because it would likely produce some biased parameters' estimations (Hilbe 2014), even though correlations between indicators are lower than 0.5. As it has been demonstrated, "[w]hen most skewnesses and/or kurtoses are larger in absolute value than 2.0, and correlations are large (say 0.5 and higher), distortions of ML and GLS chi-squares and standard errors are very likely, although estimates seem robust when relating to the *y* model" (Muthén and Kaplan 1985: pp. 187-188). In the cases, the *Generalised Least-Squares* (GLS) estimation method is a variant of ordinary least squares which has the advantage of not requiring distributional assumptions but still allowing for probabilistic inference about model fit (Hilbe 2014). Nonetheless, this estimator suffers if distributions are highly kurtotic. In this case, *Browne's Asymptotically Distribution-Free* (ADF) criterion represents a suitable alternative (Browne 1984). In fact, at the time of testing the null hypothesis that the model is correct, the

Bollen-Stine bootstrap test for goodness-of-fit measures (Bollen and Stine 1992) can be used to test overall model fit.

In terms of model fit measurements, a structural model is said to be nonrecursive if there are paths in both directions between one or more pairs of endogenous variables. In other words, this means that it is possible to start at any one of the variables in the subset, and, by following a path of single-headed arrows, return to the original variable while never leaving the subset. The presence of a feedback loop in nonrecursive models can generate problems at the time of estimating parameters related to the endogenous variables. For this reason, it is important to assess if the model is stable (Fox 1980). The stability index can be used for this purpose (Bentler and Freeman 1983). Several indexes can be computed to assess overall model fit. A *Comparative Fit Index* (CFI) close to 1 signals a very good fit (Bentler 1990); the same happens in the case of the *Parsimony Adjustment Fit Index* (*PCFI)* (Mulaik, James et al. 1989). *CMIN* indicates the minimum value on the discrepancy function, while *DF* indicates the degree of freedom. When the ratio of CMIN divided by DF is relatively close to 1, it indicates an acceptable fit between the hypothetical model and the sample data (Marsh and Hocevar 1985). A *Root Mean Square Error of Approximation* (RMSEA) value relatively close to 0 shows also an acceptable fit (Steiger and Lind 1980, Browne and Cudeck 1993). The *Akaike Information Criterion* (AIC) (Akaike 1973, Akaike 1987) and the *Bayes Information Criterion* (BIC), show good fit when they show decreasing values at the time of comparing alternative versions of the model (Raftery 1993).

5.13.3 Test of indirect effects

To assess intervening-variables effects, the statistical package PROCESS for IBM SPSS, written by Andrew F. Hayes (Field 2013, Hayes 2013), can be used to test the full list of indirect effects. In case data meet the assumption that the sampling distribution of the indirect effects is normal, The Sobel test could be used (Hayes 2009); this test is a sort of a specialised *t* test which determines whether the mediation effect is statistically significant if the reduction it determines in the effect of the independent variable, after including the mediator in the model, represents a

significant reduction. If the distribution of the indirect effects is not-normally distributed, the bootstrapping method can be used instead. In this case the inference is based on a distribution-free estimate of the indirect effect itself, usually based on 1000 bootstrap samples. This procedure yields a bootstrap confidence interval: if zero is not between the lower and upper bound, then the researcher can claim with relative confidence that the indirect effect is not zero (Hayes 2009).

5.13.4 Additional exploratory analyses

Descriptive statistics will be used to present information related to data breach occurrence frequency and reactions to the proposed General Data Protection Regulation. Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) will be used to explore reasons behind information security investments. This analysis will reflect the perceptions and understanding of the people working in the enterprise. The Mean Square Contingency coefficient, also known as phi coefficient, will be used to test the degree of association between various privacy and security measures which can be adopted by an enterprise to protect data privacy.

5.14 Conclusion

The present study investigates organisational data protection practices by exploring the effects of factors such as the impact of data protection laws, organisational privacy culture and big data analytics on organisational information management decisions. In trying to answer the research question, propositions clarifying the relationship between these factors have been identified in chapter four. Within this chapter, a survey design methodology was chosen as an appropriate data gathering method to address the phenomenon under study. Operationalisation of constructs by means of multi-item scales has been discussed. The researcher has adopted a critical realist perspective as the epistemological perspective in this study. In doing so, the researcher recognises the fundamental role played by theory and research method triangulation in social sciences in general, and business research in particular. Although the study relies on the use of

statistical methods to test specific hypotheses by means of path analysis, the researcher acknowledges the profound subjectivity and discretionality of the research process and the need of constantly integrating and comparing quantitative research results with considerations based on qualitative data, such as practitioners' experience, case studies, and stakeholders' and business commentators' opinions. For this reason, while results will be presented in chapter six, a separate chapter will be devoted to the discussion of their implications. In doing so, Chapter Seven will rely not only on academic studies but also on legal opinions, white papers and other sources.

CHAPTER SIX

Analysis and Presentation of Results

6.1 Introduction

This chapter addresses issues related to data structure and features, and data analysis. The chapter initially presents the data, followed by considerations related to data quality, participants' selection bias and non-response rate; then it moves to the test of the hypotheses identified in chapter four. Within this second part the chapter presents the assessment of data-model fit, and the identification and estimation of model parameters and the presentation of results (Mueller and Hancock 2008). Implications of results and limitations of the study are discussed in Chapter Seven.

6.2 Data collection strategy, potential bias and generalizability of results

6.2.1 Respondents characteristics

In total 442 people clicked on the electronic link to the survey and replied to some or all questions. The survey was anonymous; respondents were not asked to provide identifiable information such as their company's name, only information on their job title, education and number of working experience. Although, on one hand, the promise of anonymity, as well as the prospect of receiving the study report, increased the chance of participating in the study (Johnson and Shipps 2013), on the other hand, the researcher could not verify the level of accuracy of responses nor the identity of respondents. However, the survey was not accessible to the general public as information on how to participate were distributed only through specific channels, such as specialised press or professional associations. In addition, data have been analysed to ensure that no unexpected distortions, caused by drop-out rate or other factors, exist.

6.2.2 Survey completion rate and dropouts

The survey instrument was relatively long and 26% of participants quit the survey after answering question number 19, while 14% of participants quit after question number 26. The outcome was a total of 206 records almost fully completed. Out of them, the researcher erased 11 records because they showed multiple missing values across several variables of interest included in the model. As reported in table 31, 47% of surveys were almost fully completed, while 40% of surveys were only partially completed. 13% of potential respondents left the website before answering any questions. As a result, the number of usable surveys was 383, while the number of cases used in running multivariate analyses was 206, and in the case of the path analysis model, it was 195.

Table 31. Percentage of completed, partially completed and started-only surveys from each survey distribution channel

	ICO2	ELITE	ICO1	IAPP	APEP-ATI	EDPS	Insid eOR	PL&B	PILOTS	N
Survey started only	17%	6%	20%	12%	8%	5%	0%	0%	0%	59
Survey completed (until Q19)	29%	15%	24%	35%	27%	47%	0%	50%	0%	114
Survey completed (until Q25)	17%	16%	15%	9%	4%	5%	0%	0%	0%	63
Survey entirely completed (Q60)	37%	63%	41%	44%	62%	42%	100%	50%	100%	206
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	
N	195	108	54	34	26	19	1	2	3	442

The questionnaire was also relatively demanding, as it contained 34 questions. Probably for this reason, some respondents, who started answering questions, quit at some point. Drop-outs were divided into three groups: (1) those that opened the page but did not answer any questions; (2) those who answered questions up to question 19; (3) and those who answered questions up to question 25. The rest of the participants went through the entire questionnaire. A *progress bar* was visible at the end of each page of the survey to help respondents see how far along they were in the survey and feel in control of the time they were allocating to filling in the questionnaire; this option was meant to improve participation rate (Johnson and Shipps 2013).

The average time spent to fill in the various sections of the questionnaire gives us some insights into the potential level of fatigue or conflicting priorities which could have pushed respondents to stop answering questions. Table 32 presents the time each group spent on the questionnaire. This information was automatically recorded by Qualtrics. Some cases were excluded (see column 'number of outliers') because the reported time was unusually high and probably indicated that the webpage with the electronic survey was abandoned by the respondents. 24 minutes on average were necessary to complete the survey. There were no substantial differences in terms of completion time across survey distribution channels (see table 33).

Table 32. Survey completion average duration

Completion rate	*Total no of cases*	*No of outliers*	*No of cases considered*	*Average time spent to fill in the survey (min)*
Survey started only	59	7	52	1,44
Survey completed (until Q19)	114	10	104	4,30
Survey completed (until Q25)	63	6	57	13,15
Survey entirely completed	206	7	199	24,21

Table 33. Survey completion average duration by distribution channel – survey entirely completed

Survey entirely completed			
Distribution channel	*Average time (min)*	*No of cases*	*No of outliers*
APEP-ATI	25,03	16	0
EDPS	21,53	8	1
ELITE	26,28	68	2
IAPP	22,33	15	1
ICO1	17,58	22	0
ICO2	22,53	72	4

In terms of difference in the composition of each group of drop-outs, no significant variations have been found between the three groups with respect to organisational workforce (fig. 18), office location (fig. 19), or type of internet browser used (tab. 34). In figures 18 and 19 each concentric ring represents the specific group of people who fully completed the survey, or who completed the survey up until question 19 or up until question 25. In table 34, the type of browser was used as a proxy to check whether survey web pages were loading slower or crashing, ultimately encouraging an individual to not respond to the survey or to drop out of it (Tuten, Urban et al. 2002).

Figure 18. Workforce distribution by group of drop-outs

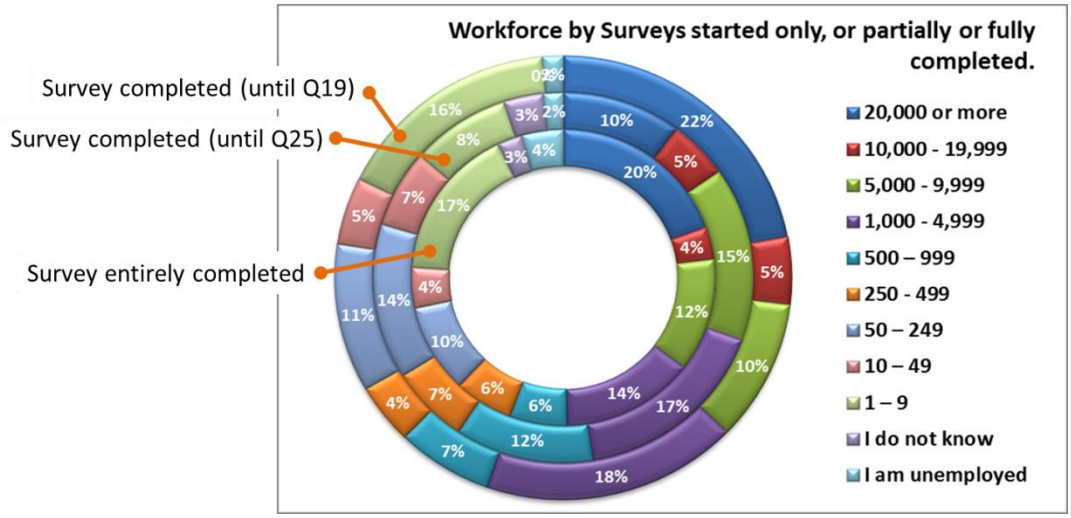


Figure 19. Office location distribution by group of drop-outs

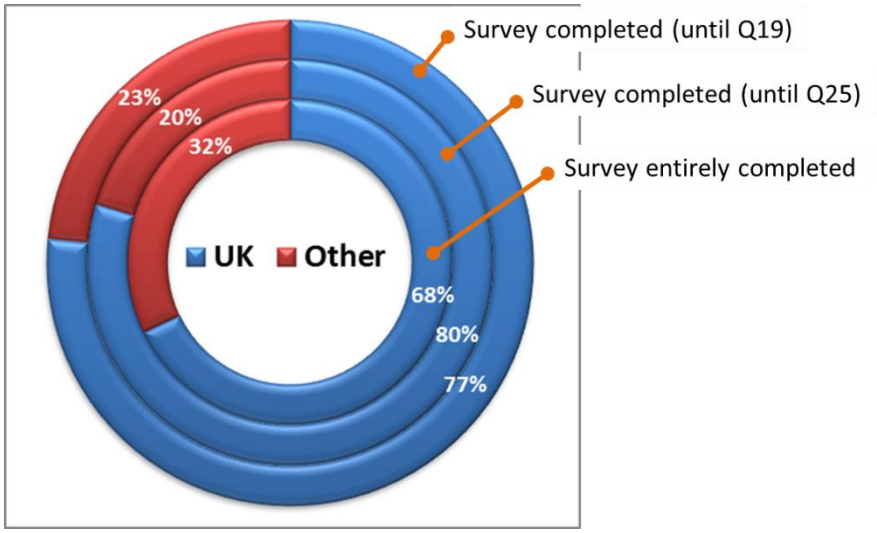


Table 34. Percentage of surveys completed by type of Internet browser

		Surveys started only, or partially or fully completed				Total
		Survey started only	*Survey completed (until Q19)*	*Survey completed (until Q25)*	*Survey entirely completed*	
Browser Meta Info-Browser	*Chrome*	20%	21%	17%	17%	82
	Mozilla Firefox	12%	12%	13%	20%	70
	Microsoft Internet Explorer (MSIE)	59%	59%	57%	52%	245
	Safari	8%	4%	6%	9%	19
	NA	0%	4%	6%	2%	12
Total (%)		100%	100%	100%	100%	
Total		59	114	63	206	442

6.2.3 Effects of survey distribution channel on responses

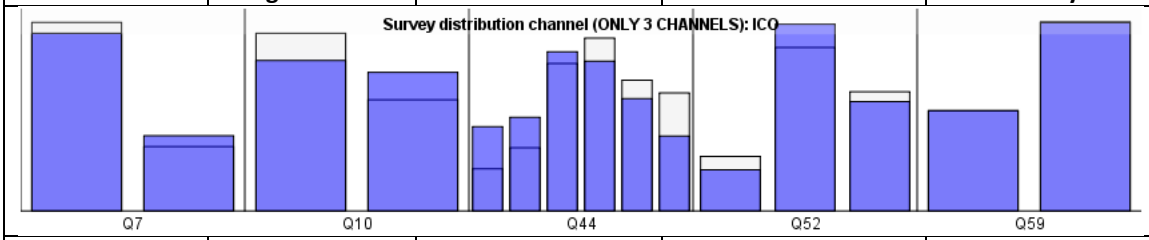
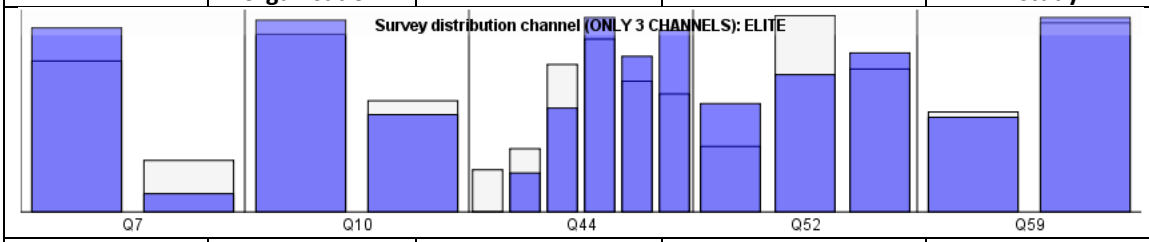
This section explores the nature of the data gathered by paying special attention to the variability caused by participants' selection bias, non-response rate, and reliability of responses. Five variables were selected and their distribution across survey distribution channel analysed. These variables are: headquarters' location (UK-based or not); business type (for profit vs. nonprofit); reported sectorial data breach frequency; organisational proactivity toward the reform of the 1995 Data Protection Directive; respondents' willingness to participate in the follow-up of the study. Each dependent variable was measured at ordinal level, while the independent variable consisted of three categorical, independent groups, each of them representing a major survey distribution channel (i.e. the ICO's newsletter; ELITE's member lists; or other channels).

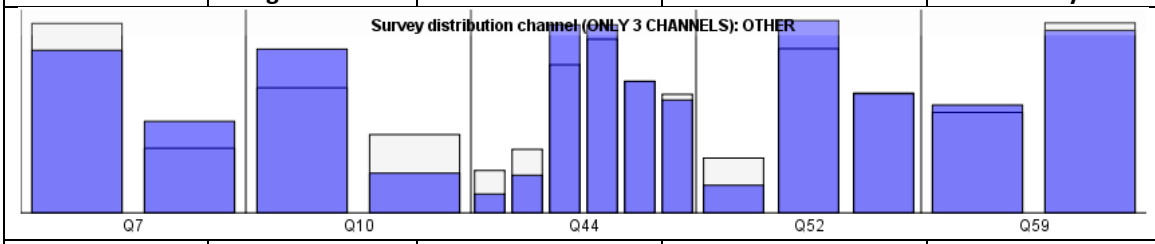
As showed in table 35 and demonstrated by the results of the Kruskal-Wallis H test reported at the bottom of the table, the distribution of responses do not significantly differ across distribution channels in the case of two out of five variables. These two variables are organisational proactivity toward the reform of the 1995 Data Protection Directive and respondents' willingness to participate in the follow-up of the study. In contrast, the distribution channel seems to be related to the probability of working for a British or for a nonprofit organisation, and to the probability of reporting a certain data breach frequency.

As expected, and as illustrated in the graphs included in table 35, members of the British Computer Society Effective Leadership in IT (ELITE) Group were mostly based in the UK (5a), while readers of the IAPP blog or members of Spanish ATI and APEP were mostly based outside the UK (6a). Readers of the ICO's newsletter mostly work for non-profit organisations (4b), while all other respondents work mostly for profit organisations (5b; 6b). Readers of the ICO newsletter (4c) and members of IAPP and similar groups (6c) reported a higher number of data breaches than members of the ELITE group (5c); it might be due to their position within the organisation, probably the legal or information security department, which increases their chances of being informed about this kind of events. People were equally interested, or not interested, in

participating in a follow-up to the study (4e; 5e; 6e), no matter from where they had received the invitation to participate originally. There is also no significant difference across distribution channel in terms of the propensity of organisations to start planning for the proposed General Data Protection Regulation (4d; 5d; 6d).

Table 35. Kruskal-Wallis equality-of-populations rank test for testing distribution variability across distribution channels.

SURVEY DISTRIBUTION CHANNEL: **ICO's newsletters**				
4a. Office based in the UK	**4b. For profit/Non-profit organisation**	**4c. Data breach frequency**	**4d. Planning or not for the GDPR**	**4e. Interested in participating in the follow-up of the study**
				
(Yes; No)	*(Yes; No)*	*(Incidents may occur on a daily, weekly, monthly, yearly basis, never heard of any, or DK)*	*(DK; Yes; No)*	*(Yes; No)*
SURVEY DISTRIBUTION CHANNEL: **ELITE's members**				
5a. Office based in the UK	**5b. For profit/Non-profit organisation**	**5c. Data breach frequency**	**5d. Planning or not for the GDPR**	**5e. Interested in participating in the follow-up of the study**
				
(Yes; No)	*(Yes; No)*	*(Incidents may occur on a daily, weekly, monthly, yearly basis, never heard of any, or DK)*	*(DK; Yes; No)*	*(Yes; No)*

SURVEY DISTRIBUTION CHANNEL: **OTHER (IAPP; APEP; Inside OR; ATI; PL&B)**				
6a. Office based in the UK	**6b. For profit/Non-profit organisation**	**6c. Data breach frequency**	**6d. Planning or not for the GDPR**	**6e. Interested in participating in the follow-up of the study**
				
(Yes; No)	*(Yes; No)*	*(Incidents may occur on a daily, weekly, monthly, yearly basis, never heard of any, or DK)*	*(DK; Yes; No)*	*(Yes; No)*
Hypothesis test results				
Office based in the UK	**For profit/Non-profit organisation**	**Data breach frequency**	**Planning or not for the GDPR**	**Interested in participating in the follow-up of the study**
chi-squared = 11.669 with 2 d.f. probability = 0.0029	chi-squared = 13.776 with 2 d.f. probability = 0.0010	chi-squared = 13.061 with 2 d.f. probability = 0.0015	chi-squared = 0.492 with 2 d.f. probability = 0.7818	chi-squared = 0.126 with 2 d.f. probability = 0.9391

6.3 Answering the research questions with path analysis

In Chapter Four eleven propositions and several corresponding hypotheses, describing the way some key factors were expected to influence the dependent variables, were identified. In this section hypotheses are tested by means of path analysis. 8 variables were used in the path analysis model with a dataset of 195 cases. Almost 25 cases were available for each variable inserted in the model.

6.3.1 Estimation method

In selecting the estimation method some considerations related to the multivariate sample distribution are necessary. Data here presented are clearly not-normally distributed: as displayed in table 36, some critical values exceed +/- 2.00, which indicates statistically significant degrees of non-normality. The Browne's Asymptotically Distribution-Free (ADF) criterion (Browne 1984) has been used to estimate model parameters. When testing the null hypothesis that the model is

correct, the Bollen-Stine bootstrap test for goodness-of-fit measures (Bollen and Stine 1992) rejects the model if the GLS estimator is used (0.045), while it retains the model if the ADF estimator is adopted (p = 0.075). A more exhaustive discussion of data characteristics, impossibility of meeting normal theory assumptions, and adoption of distribution-free statistics is included in the Statistical Appendix.

Table 36. Assessment of normality

Standardised Variables	Value Min	Value Max	Original Variables	Value Min	Value Max	Skewness	Critical Ratio	Kurtosis	Critical Ratio
ZSOPH_N	-2.65	2.51	SOPH_N	-7	0	-0.06	-0.33	-0.50	-1.42
ZREG_N	-2.48	1.37	REG_N	-7	0	-0.71	-4.07	-0.07	-0.21
ZFUNCT_P	-1.89	2.21	FUNCT_P	0	100	-0.12	-0.70	-0.53	-1.51
ZDPOOL_P	-1.21	3.35	DPOOL_P	0	100	1.01	5.78	0.79	2.24
ZDVEIL_N	-1.39	1.63	DVEIL_N	-7	0	0.12	0.67	-1.22	-3.47
ZPRV_N	-2.66	1.41	PRV_N	-7	0	-0.71	-4.02	-0.15	-0.43
ZDSR_N	-2.59	1.56	DSR_N	-7	0	-0.33	-1.91	-0.35	-0.99
ZDPP_N	-2.25	1.84	DPP_N	-6	0	0.07	0.38	-0.47	-1.33
Multivariate								7.02	3.88

6.3.2 Model fit measurements

The structural model designed is nonrecursive. The stability index for the subset composed of the variables DPOOL, DVEIL and PRV is 0.056, which is far less than 1, meaning that the system of linear equations associated with the model can be considered ‘stable’.

Measurements of model fit are reported in table 37. The *Root Mean Square Error of Approximation* (RMSEA) is relatively close to 0 showing an acceptable fit (Steiger and Lind 1980, Browne and Cudeck 1993). The ratio of CMIN divided by DF is relatively close to 1 which is indicative of an acceptable fit between the hypothetical model and the sample data (Marsh and Hocevar 1985). The *Comparative Fit Index* (CFI) is close to 1 signalling a very good fit (Bentler 1990). The Akaike information criterion (AIC) (Akaike 1973, Akaike 1987) and the Bayes Information Criterion (BIC), show good fit – i.e. decreasing values – if compared with alternative versions of the model (Raftery 1993).

Table 37. Model fit summary

NPAR	***CMIN***	***DF***	***P***	***CMIN/DF***	***CFI***
33	8.059	3	0.045	2.686	0.974
Stability index	***RMSEA***	***AIC***	***BIC***	**PRATIO**	**PCFI**
0.052	0.093	74.06	182.07	0.107	0.104

6.3.3 Test of hypotheses

Standardised regression weights are reported in figure 20 and in table 38. Whether each hypothesis has been rejected or not rejected is reported in the last column of table 38, which summarises model results. The path analysis provides support for almost all hypotheses. Only two hypotheses were rejected. These are: HB11, which states that the data protection regulatory regime would limit data collection; and HE32, which states that the more organisations rely on analytics across business functions, the more likely it will be that they also rely on targeted analytics. Hypothesis HE11 seems to be rejected in the path analysis model because of the significance level chosen (Alpha = 0.05; p-value = 0,056). In the next section, the analysis of indirect effects shows the presence of a positive relationship between the organisational privacy culture and the use of targeted analytics, and that the organisational privacy culture mediates the relationship between analytical sophistication and the use of dataveillance as targeted analytics.

Table 38. Regression weights: Asymptotic Distribution Free (ADF) estimates

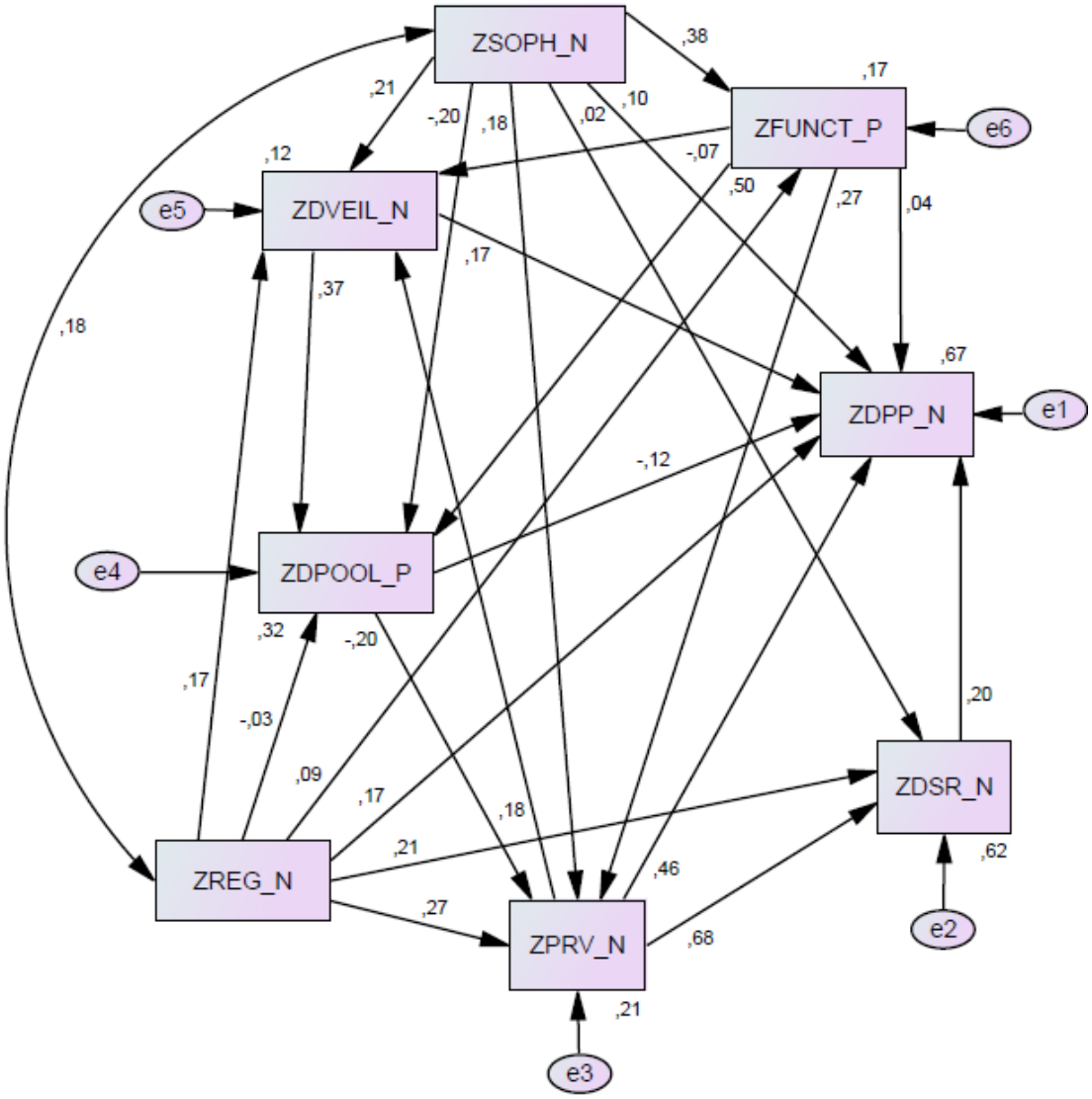
No	Hp	DP		IV	Estimate	S.E.	Sig.
1	***H-A.11***	DPP	⇦	REG	0.171	(0.049)	***
2	***H-A.12***	DSR	⇦	REG	0.212	(0.053)	***
3	***H-A.13***	PRV	⇦	REG	0.263	(0.072)	***
4	***H-A.21***	DPP	⇦	PRV	0.460	(0.082)	***
5	***H-A.22***	DSR	⇦	PRV	0.684	(0.063)	***
6	***H-A.23***	DPP	⇦	DSR	0.204	(0.077)	**
7	***H-B.11***	DPOOL	⇦	REG	-0.034	(0.065)	R
8	***H-B.21***	DVEIL	⇦	REG	0.169	(0.074)	*
9	***H-B.22***	DPP	⇦	DVEIL	0.166	(0.054)	**
10	***H-C.11***	FUNCT	⇦	SOPH	0.382	(0.069)	***
11	***H-C.12***	DPOOL	⇦	FUNCT	0.494	(0.071)	***
12	***H-C.21***	DVEIL	⇦	SOPH	0.213	(0.079)	**
13	***H-C.22***	DPOOL	⇦	DVEIL	0.374	(0.067)	***
14	***H-C.31***	DPOOL	⇦	SOPH	-0.198	(0.073)	**
15	***H-D.11***	DPP	⇦	DPOOL	-0.115	(0.053)	*
16	***H-D.12***	PRV	⇦	DPOOL	-0.194	(0.083)	*
17	***H-E.11***	DVEIL	⇦	PRV	0.182	(0.095)	R
18	***H-E.21***	PRV	⇦	SOPH	0.177	(0.071)	*
19	***H-E.31***	PRV	⇦	FUNCT	0.265	(0.085)	**
20	***H-E.32***	DVEIL	⇦	FUNCT	-0.072	(0.075)	R
21		FUNCT	⇦	REG	0.091	(0.075)	R
22		DPP	⇦	SOPH	0.094	(0.052)	R
23		DSR	⇦	SOPH	0.022	(0.055)	R
24		DPP	⇦	FUNCT	0.042	(0.054)	R

KEY

Significance level Alpha: *** 0.001; ** 0.01; * 0.05

R = Hypothesis rejected

Figure 20. Path analysis: Asymptotic Distribution Free (ADF) estimates



6.4 Testing for mediation effects

Table 35 presents the tests for the presence of indirect effects, namely the potential mediation effects identified in chapter four. While full table of results are reported in the Statistical Appendix, table 35 contains a summary of the most relevant indirect effects identified. The bootstrapping method was used to assess whether indirect effects were significantly different from zero. The Sobel test was not used since it requires the assumption that the sampling distribution of the indirect effect is normal (Hayes 2009).

Since zero is not between the lower and upper bound of the bootstrap confidence interval, the researcher can claim with relative confidence that there is an indirect effect (Hayes 2009). This analysis provides support for all hypothesised mediation effects but one, which is M-E.3.

Table 39. Bootstrapping analysis of indirect effects

	Mediation	Point estimate	95% Confidence Interval		Outcome
			Lower	Upper	
M-A.2	PRV ➔ DSR ➔ DPP	0.211	0.103	0.335	Partial mediation
M-B.2	REG ➔ DVEIL ➔ DPP	0.066	0.029	0.126	Partial mediation
M-C.1	SOPH ➔ FUNC ➔ DPOOL	0.164	0.096	0.242	Full mediation
M-C.2	SOPH ➔ DVEIL ➔ DPOOL	0.077	0.032	0.150	Full mediation
M-E.1	SOPH ➔ PRV ➔ DVEIL	0.047	0.002	0.103	Partial mediation
M-E.2	SOPH ➔ FUNC ➔ PRV	0.076	0.020	0.153	Partial mediation
M-E.3	FUNC ➔ PRV ➔ DVEIL	0.059	0.019	0.123	No effect of FUNC on DVEIL

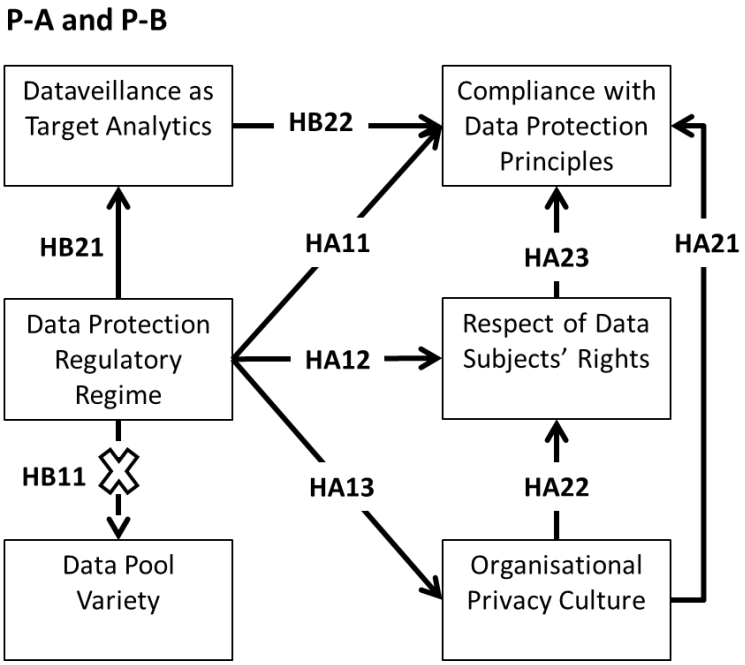
6.5 Summary of results

Based on the results of the path analysis and on the test of indirect effects, this section presents a summary of results with a clear reference to the research questions. Results will then be discussed in Chapter Seven.

Answering Research Question One: *How does the data protection regulatory regime influence enterprise data protection and data management decisions?*

Propositions A and B answer Question One. To provide a complete overview of these propositions, figure 21 shows relationships between constructs. The presence and direction of association is signalled by an arrow. Dot lines represent negative relationships.

Figure 21. Answering Research Question One: Propositions A and B and corresponding hypotheses



According to the results of the path analysis model and the test of intervening-variables effects, the researcher draws the following conclusions. The less permissive and more reliable the data protection regulatory regime, the more likely it will be that organisations develop an internal privacy culture (HA13), respect data subjects' rights (HA12) and comply with data protection principles (HA11). Furthermore, the more organisations foster their internal privacy cultures, the more likely it will be that they respect data subjects' rights (HA22) and comply with data protection principles (HA21). As suggested by surveillance scholars and reported in section 4.6.2, the current privacy regime is compatible with the deployment and application of targeted analytics (HE11). For organisations which want to use targeted analytics privacy is a very important topic: organisations tend to acknowledge that initiatives which target individuals with personalised offers need to be designed in a privacy-sensitive way. Otherwise companies know they face the risk of suffering both prosecution and consumer backlash. Thus, organisations try to use targeted analytics in a way compliant with data protection principles (HB22). To pursue this objective, employees receive privacy training, and different measurements to protect data from abuse are adopted.

Thus, the use of big data analytics seems to be compatible with the development of an organisational privacy culture. Additionally, and in contrast with what suggested elsewhere (Lester 2001), the study finds no support for the hypothesis which states that the presence of a strict and reliable data protection regulatory regime would create barriers to data reuse, collection and analysis (HB11). Based on the results of the test of hypotheses HB21 and HB11, this study does not find support in favour of the argument, presented in section 4.5, which states that data protection law would prevent innovation or disrupt business operations. The current European legal privacy regime seems to be compatible with technological developments such as big data analytics.

The lack of association between the regulatory regime and the amount and variety of data processed by the organisation find further support in the analysis included in table 40. The researcher has not found any sign of association between the composite score Data Protection Regulatory Regime and the variable which measures the amount of data processed by the organisation.

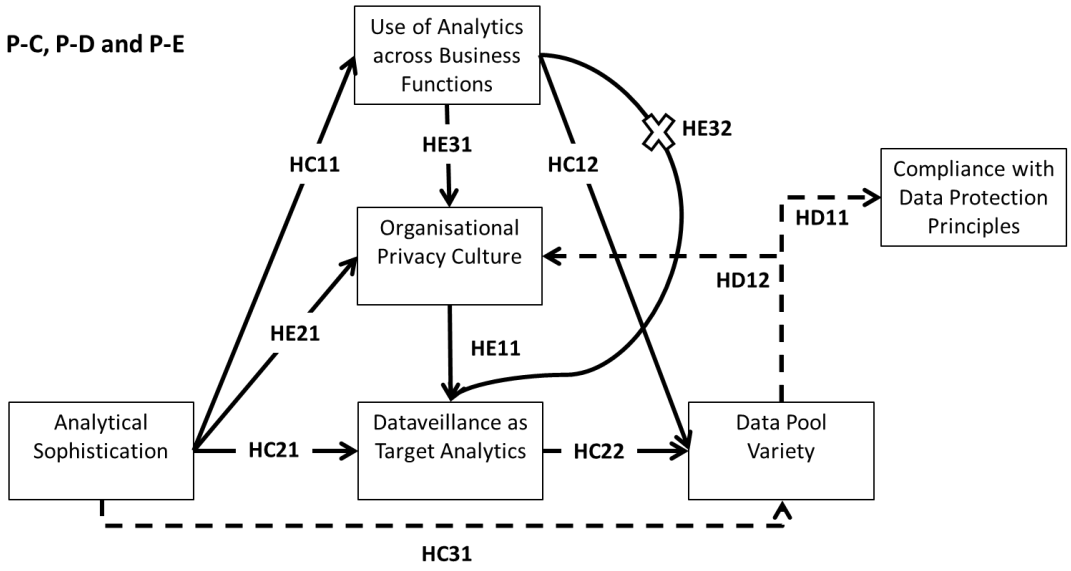
Table 40. Relationship between the Privacy Regulatory Regime and the Amount of data processed by the organisation expressed in terabytes

Nonparametric Association Tests	
Spearman's Rank Order	Kendall's rank correlation
spearman Q25 REG_N, star(0.05) Number of obs. = 188 Spearman's rho = -0.0152 Test of Ho: Q25 and REG_N are independent Prob > |t| = 0.8356	ktau Q25 REG_N, star(0.05) Number of obs. = 188 Kendall's tau-a = -0.0102 Kendall's tau-b = -0.0118 Kendall's score = -180 SE of score = 841.706 (corrected for ties) Test of Ho: Q25 and REG_N are independent Prob > |z| = 0.8316 (continuity corrected)

Answering Research Question Two: *How does the level of analytical sophistication an organisation has achieved influence enterprise data protection and data management decisions?*

Propositions C, D and E answer Question Two. Figure 22 offers an overview of these propositions and of the relationships between constructs.

Figure 22. Answering Research Question Two: Propositions C, D and E and corresponding hypotheses



According to the results of the path analysis model and the test of indirect effects, the researcher draws the following conclusions. As suggested by the literature on analytical competitors discussed in sections 2.5 and 4.6, organisations which are analytically sophisticated are more likely to rely on targeted analytics (HC21), and on other types of analytical tools to achieve a wide variety of objectives (HC11). However, there is no direct relationship between relying on targeted analytics and the use of analytics across different organisational units (HE32).

While these findings confirm results from previous studies, hypothesis HE21 supports the view that analytical competitors tend to be privacy champions (Davenport, Harris et al. 2010). As discussed in section 4.6.1, the more organisations are analytically sophisticated, the more likely it will be that they develop an internal privacy culture (HE21) compatible with the use of targeted analytics (HE11). Additionally, the construct Privacy Culture (PRV) partially mediates the relationship between the constructs Analytical Sophistication and Dataveillance (DVEIL) interpreted as Targeted Analytics, though the relationship between PRV and DVEIL (HE11) is relatively weak and it was not detected in the path analysis model.

Regarding the relationship between big data and data protection, it seems that the more analytically sophisticated organisations are, the less likely it will be that they indiscriminately

collect and store a large variety of data (HC31). Considerations related to data minimisation and data quality could play a role in persuading organisations to limit data accumulation. The more analytics is also used by different departments, or units, within the organisation, the more likely it will be that all members of staff receive privacy training and that privacy-preserving procedures are adopted by the entire enterprise (HE31).

Although, at first glance, big data analytics and the current privacy regime appears to be almost perfectly compatible, important frictions still exist between the logic of big data and the logic of data protection. Conflicts emerge on data collection, fusion and retention. The more organisations process and analyse a large variety of data, the less likely it will be that they develop a privacy culture (HD12) and comply with data protection principles (HD11). The demand for data fusion and data accumulation is also driven by analytics. Organisations which rely on targeted analytics (HC22), and on all kind of analytical tools within the organisations (HC12), are more likely to collect and process a large amount of data in different formats.

Therefore, contradictions between the logic of big data and the logic of data protection emerge around the issue of data collection. The more analytically sophisticated organisations become, the more they invest in analytics across business functions, the more they need data. The demand for analysis-based answers drives the demand for data acquisition, accumulation, and matching. Although efficient data management systems would tend to adopt a "data-minimisation" logic and reduce the amount of data retained, the demand for increasingly granular data to personalise offers, or forecast demand, push toward the integration of different data sources and the treatment of personal information. As pointed out by the literature on digital surveillance (Clarke 1988, Gandy 1993, Lyon 1993, Andrejevic 2009, Degli Esposti 2014), targeted analytics, interpreted as a manifestation of dataveillance, reinforces data accumulation and contributes to extending and amplifying digital surveillance through the proliferation of data gathering instruments and database integration.

Nevertheless, analytically sophisticated organisations which use analytics to pursue different objectives are aware that they have to deal with information privacy issues, educate staff and adopt privacy-preserving procedures. For this reason, they invest in both in privacy awareness campaigns and information security training, as well as in targeted analytics solutions. Companies have probably learned the importance of taking into account customers' reactions at the time of applying sophisticated analytical tools. Developing a privacy culture seems to become a precondition before starting to adopt analytics across business units.

While the implications of these results will be discussed in the next chapter, the next section pays more attention to issues related to information security investment decisions and the adoption of privacy-preserving procedures. As considerations related to information security are strictly intertwined with data protection considerations, the following section hopes to further contribute to our understanding of organisations' data management decisions related to data privacy and information security.

6.6 Additional analyses

6.6.1 Exploring information security investment decisions

In Chapter Three, section 3.8, the topic of which elements drive investments in information security has been discussed. Although studies on the economics of information security tend to focus on market reactions to security investments (Spanos and Angelis 2016), in this section the researcher will try to explore the rationale behind information security investments from the perspective of survey respondents. Table 41 presents descriptive statistics of seven questions asked in the survey about potential reasons for investing in information security. Kernel density distributions for these variables are displayed in the Statistical Appendix.

In order to explore potential logics underlying information security investments, factor analysis has been used. By means of Exploratory Factor Analysis, and Confirmatory Factor Analysis with Principal Axis Factoring Method (KMO = 0.856), the researcher has identified two factors capable

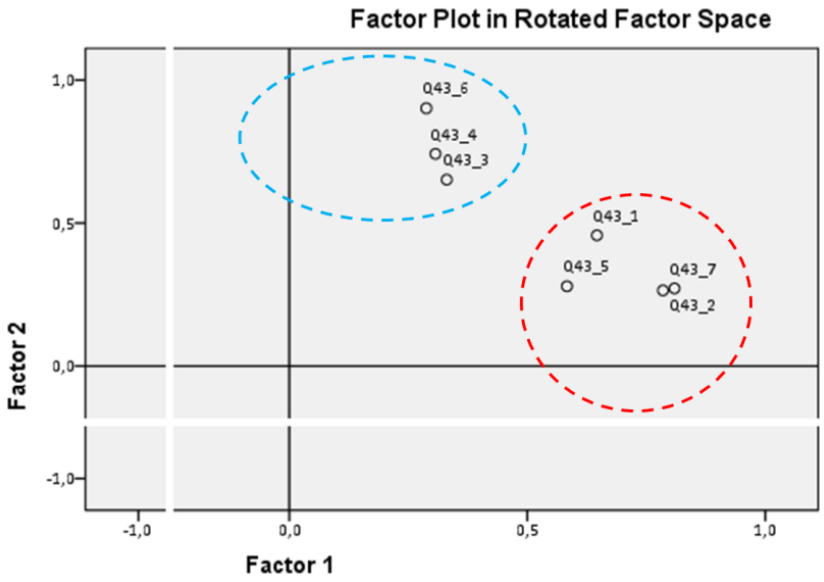
of summarising the information contained in these seven questions. The two factors extracted explained 74% of the total variance. The Equamax rotation method, which minimises both the number of variables that load highly on a factor, similarly to the Varimax method, and the number of factors needed to explain a variable, as does the Quartimax method, was used to increase readability of results (see figure 23 and table 42).

Two fundamental logics seem to emerge from the analysis of the data. On one hand, Factor 1 identifies those organisations which invest in information security mainly to: (a) react to previous security problems; (b) manage the risk of economic loss; (c) avoid costly enforcement action by regulators, and (d) reduce the risk of paying high litigation costs. On the other hand, Factor 2 identifies those organisations whose investments in information security reflect (e) high industry information security standards, (f) high service or product quality, and (g) concerns for the potential reputational risks caused by a data breach.

Table 41. Motives behind investing in InfoSec: Descriptive statistics

Variable Label	Obs.	Mean	Std. Dev.	Min	Max
Q43_1 - To manage the risk of economic loss	154	60.655	29.731	0	100
Q43_2 - To manage the risk of high litigation costs	151	53.199	33.494	0	100
Q43_3 - To manage reputational risks	161	73.385	27.484	0	100
Q43_4 - To improve service/product quality	155	57.993	32.338	0	100
Q43_5 - To react to previous security problems	146	51.452	31.084	0	100
Q43_6 - To reflect high industry information security standards.	156	62.756	30.642	0	100
Q43_7 - To avoid costly enforcement action by regulators	157	59.949	32.553	0	100
Q43_8 - I do not know	12	14.833	22.417	0	66

Figure 23. InfoSec Investments: Factor plot in Equamax space



Extraction Method: Principal Axis Factoring. Rotation Method: Equamax with Kaiser Normalization.

Table 42. Motives behind investing in InfoSec: Rotated factor matrix

Rotated Factor Matrix	Factor 1	Factor 2
Q43_1 - To manage the risk of economic loss	**.646**	.457
Q43_2 - To manage the risk of high litigation costs	**.785**	.264
Q43_3 - To manage reputational risks	.331	**.652**
Q43_4 - To improve service/product quality	.307	**.742**
Q43_5 - To react to previous security problems	**.583**	.279
Q43_6 - To reflect high industry information security standards.	.288	**.901**
Q43_7 - To avoid costly enforcement action by regulators	**.809**	.271

Information security can be perceived within an organisation as an additional burden or as a dimension of quality. Events such as cyber security incidents, the loss, or theft, of confidential information, regulatory initiatives which require organisations to disclose data breaches, play a role in changing the way the topic of information security is treated within an organisation (BVCA 2015). As a consequence, the next section explores the data privacy and cyber security risks organisations may face.

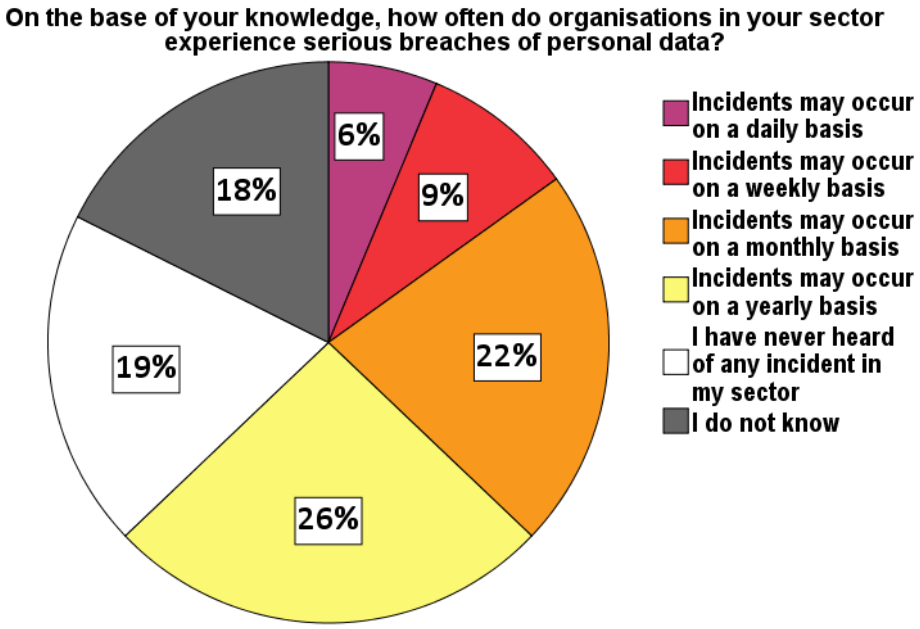
6.6.2 Frequency of data breaches

With increased information security breaches worldwide, there has been a pressing need to keep organisations' information systems secure. Yet "the management of information security is a

much deeper and more political problem than is usually realized; solutions are likely to be subtle and partial, while many simplistic technical approaches are bound to fail" (Anderson 2001: p. 6). Results of a survey based on 1,125 respondents shows that 81% of large organisations and 60% of small businesses had a security breach in 2013-14 (BIS 2014).

Since most information security management decisions taken by organisations are based on past security incidents, it is worth exploring whether perceptions of the frequency of occurrence of these events influence the level of resources employed to protect information. Because of the sensitivity of the issue, the question included in the survey asked generically for the frequency of occurrence of data breaches in the sector where the organisation operates (see figure 24). There is currently very little transparency in Europe on data breaches. In contrast with the US, European organisations in general do not have the obligation to report incidents related to compromised personal information.

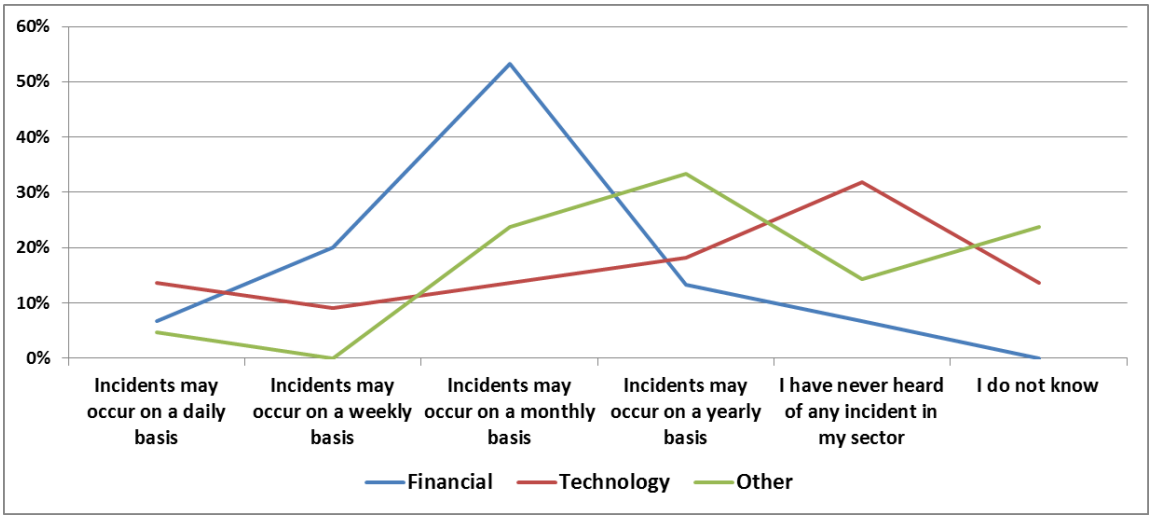
Figure 24. Data breach frequency of occurrence (n = 159)



Most respondents did not know anything about data breaches in the sector their organisations operate. By looking at the type of responses given by professionals working in for profit organisations we can see that professionals working in the financial sector knew incidents can occur on a monthly basis, while professionals working in technology companies estimated that

incidents may occur on a yearly basis. Professionals working in all other sectors were largely uninformed about this issue.

Figure 25. Data breach frequency of occurrence by sector: For profit firms (n = 58)



With respect to the type of incident or event which could cause a data breach (see table 43), unintended disclosure of sensitive information was considered a likely cause by 18% of respondents. These results are consistent with previous studies in the area (Degli-Esposti 2012).

Table 43. Common causes of data breaches

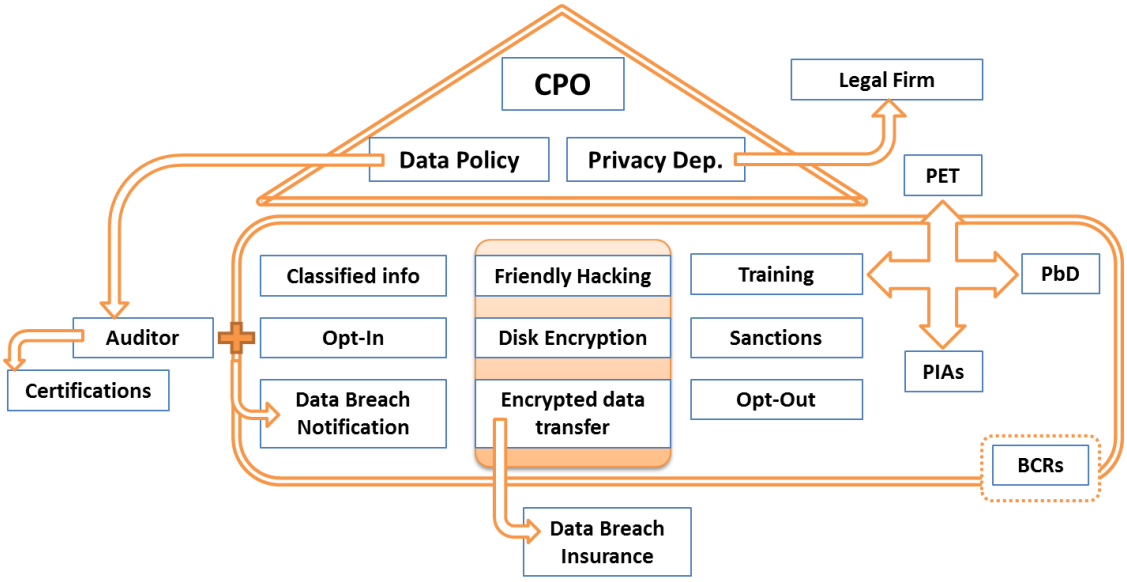
In your opinion, what most commonly causes data breaches? (n = 206)	Freq.	Percent.
1. Unintended disclosure (e.g. sensitive information posted publicly on a website or sent to the wrong party via email).	37	18%
2. Insiders (someone with legitimate access, such as an employee or contractor who intentionally breaches information).	25	12%
3. Lost, discarded or stolen portable device (e.g. laptop, PDA, smart-phones, USB, CDs...).	24	12%
4. Lost, discarded or stolen non-electronic records (e.g. paper documents).	21	10%
5. Hacking, malwares or spywares.	16	8%
6. Lost, discarded or stolen stationary electronic device (e.g. desktop computers, servers...).	15	7%
7. Payment Card Fraud (e.g. skimming devices at point-of-service terminals).	4	2%
8. Other: Google search	1	0.5%
9. Other: people / complacency	1	0.5%
10. I do not know.	2	1%

6.6.3 Relationships between privacy and security measures

As discussed in section 3.9, investing in information security and data privacy entails a number of actions which can require human, technical, legal and organisational resources. This section explores the kind of measures actually adopted by organisations and the relationships among them (Questions Q53_1-20: "Which privacy or security safeguard has your organisation already adopted? *More than one answer allowed")*. Since these questions produced dichotomous variables, the mean square contingency coefficient, known as Phi, was used to investigate bivariate relationships between these categorical variables (Guilford 1941). Phi Coefficients and descriptive statistics, namely the frequency of adoption of each data privacy or security measure, have been included in the Statistical Appendix.

Figure 26 represents a graphical representation of the interrelationships between the measures considered. According to the results of the nonparametric association test performed, the following considerations can be made.

Figure 26. Interrelationships between privacy and security safeguards



In general, a specific department inside the organisation, such as the compliance office or the IT department is in charge of dealing with privacy-related matters. Sometimes this unit can be led by a Chief Privacy or Data Protection Officer. According to previous studies (Shalhoub 2009), CPOs are expected to: (a) educate workforce in the fundamentals of fair information practices; (b)

observe compliance with privacy laws; (c) assist with the development of privacy impact assessment (PIAs); (d) promote privacy in conjunction with security; make privacy part of the fabric of the organisation; and (e) communicate privacy concerns and issues with top management. As presented in table 44, this study confirms the relationship between the CPO functions and most of these competences.

Table 44. *Relationship between the CPO's function and other privacy and security safeguards: Phi coefficients*

		Phi coefficient			Phi coefficient
No.	**CATEGORY**	**CPO**	**No.**	**CATEGORY**	**CPO**
1.	Data Policy	,710	10.	Auditor	,443
2.	Privacy Dept.	,623	11.	Opt-out	,432
3.	Training	,598	12.	Opt-in	,425
4.	Disk encryption	,597	13.	PbD	,399
5.	Friendly hacking	,590	14.	Certifications	,397
6.	Classified info policy	,590	15.	Notification	,394
7.	Sanctions	,589	16.	PETs	,373
8.	Encrypted data transfer	,581	17.	Legal firm	,343
9.	PIAs	,488	18.	BCRs	,284
			19.	Data breach insurance	,243

The privacy team works in three fundamental areas: (a) the development of internal data handling policies and procedures; (b) the promotion of workforce privacy training; and (c) the assessment of security system vulnerabilities and resilience. First and foremost, the privacy team ensures that data policies are adopted. These policies describe the rules controlling the integrity, security, quality, and use of data during its life-cycle and state change. They also envision special clauses for classifying information according to their sensitivity (e.g. secret; confidential; for internal use; etc.). In pursuing these objectives, the team can rely on the counsel of external legal firms specialized in information privacy.

The privacy team also organises training activities to help workforce members comply with privacy procedures. Some organisations envision sanctions for those employees who fail to comply with these procedures. These organisations are familiar with running network and application penetration and vulnerability tests (e.g. 'friendly hacking').

Organisations which rely on consent obtained through opt-in acceptance of data processing terms and conditions are more likely to give immediate notification to individuals if their data are breached, disclosed or manipulated. This type of organisation receives periodical visits from external auditors involved in the assessment of internal security procedures as part of certification programmes in the area of information security management (e.g. ISO/IEC 27002:2005).

Exercises to test human and technical vulnerabilities are also envisioned and undertaken as part of the overall data privacy strategy. Privacy Impact Assessments (PIAs) are undertaken in conjunction with employees' training activities and can lead to the adoption of opt-out consent-forms; technical measures such as full-disk encryption of physical devices like laptops or PCs and encrypted transmission of data are also adopted.

Finally, while the large majority of organisations have data policies and a team of professionals dealing with the issue of privacy, a few organisations adopt Privacy-by-design (PbD) criteria in product development or use Privacy Enhancing Technologies (PETs). Very few organisations also have data breach insurance policies or adopt Binding Corporate Rules (BCRs) to manage international data transfer.

Finally, as suggested in previous qualitative studies (Da Veiga and Martins 2015), the creation of an internal privacy and information security culture contribute to enhance workforce awareness and knowledge, which produce a consequent improvement of data privacy and security practices and procedures. As showed in table 45, the Privacy Culture index correlates with most of privacy and security measures explored in this section. Kendall's rank correlation coefficients show fairly strong association between an organisation privacy culture and the presence of data policies, a Chief Privacy Officer, workforce privacy training, PETs, PbD, reliance on opt-in consent procedures, sanctions for data mismanagement, application of Binding Corporate Rules, several security measures like data encryption or penetration tests.

Table 45. *Relationship between the composite score Privacy Culture and each privacy and security safeguards: Kendall's Rank Correlation coefficients*

		PRV			PRV
PRV	Privacy Culture composite score	0.915	PRV	Privacy Culture composite score	0.915
Q53_1	A Chief Privacy/Data Protection Officer is in charge of supervising all privacy-related issues.	**0.089***	Q53_11	Privacy Impact Assessments (PIAs) are undertaken.	0.043
Q53_2	The function of dealing with privacy-related matters is pursued by a designated department inside my organisation, for example the compliance office or the IT department, etc.	0.078	Q53_12	Counsel of a legal firm specialized in information privacy.	0.018
Q53_3	Data policies that describe the rules controlling the integrity, security, quality, and use of data during its life-cycle and state change have been adopted.	**0.103***	Q53_13	Binding Corporate Rules (BCRs) to manage international data transfer.	**0.106***
Q53_4	Specific policies for classifying information according to their sensitivity (e.g. secret; confidential; for internal use; etc.) are in place.	**0.103***	Q53_14	Periodical external auditors' assessment of internal security standards.	-0.036
Q53_5	Consent obtained through opt-in acceptance of data processing terms and conditions.	**0.105***	Q53_15	Immediate notification to individuals if their data are breached, disclosed or manipulated.	**0.087***
Q53_6	Consent obtained through opt-out acceptance of data processing terms and conditions.	0.028	Q53_16	Certified code of practice for information security management (e.g. ISO/IEC 27002:2005).	0.047
Q53_7	Employees are constantly trained to comply with privacy procedures.	**0.218***	Q53_17	Data breach insurance policy.	0.033
Q53_8	Workforce members are sanctioned if they do not comply with privacy procedures.	**0.177***	Q53_18	Full-disk encryption of physical devices like laptops or PCs.	0.130*
Q53_9	Privacy Enhancing Technologies (PETs) are in use.	**0.150***	Q53_19	Encrypted transmission of data.	**0.156***
Q53_10	Privacy-by-design (PbD) criteria are adopted in product development.	**0.127***	Q53_20	Network and application penetration and vulnerability testing (e.g. friendly hacking).	**0.094***

* Significance level Alpha 0.05

6.6.4 Reactions to the proposed General Data Protection Regulation

Regarding the provisions of the proposed new General Data Protection Regulation, discussed in section 3.10, respondents found the following provisions somewhat or highly problematic (see table 46). The right of erasure, also known as the right-to-be-forgotten, is considered problematic by the large majority of organisations, followed by the right to data portability. Half of organisations consider provisions such as the adoption of Binding Corporate Rules, Data Protection Impact Assessment, and Privacy by Design principles also problematic. Only one third of respondents consider the provisions on explicit consent or compulsory data breach notification problematic. The least problematic provision is the one which envisions the appointment of a data protection officer within the organisations supervising data processing activities.

Table 46. Percentage of respondents that consider each provision of the GDPR problematic

*All orgs. (**n = 167**)*	*Orgs working with data (**n = 53**)*	*Provisions envisioned in the proposed General Data Protection Regulation (GDPR)*
72%	68%	Data subjects will have the **right to erasure**. This will allow individuals to have all personal data that business holds on them deleted or restricted. This will include all photos and any public links to, or copies of, personal data that can be found on the Internet for example in social networks or via search engines.
66%	66%	Data subjects will have the **right to data portability**, which is a right to require a portable copy of a data subject's personal data so that they may transfer it to another data controller.
60%	55%	The regulation will apply to organisations outside the EU whenever they process personal data of individuals in the EU. Data transfer outside the EU will be possible through **Binding Corporate Rules (BCR)** or in case of authorisation given by data protection authorities. Authorisations will be valid only for two years.
53%	47%	**Data Protection Impact Assessment (PIA)** must be performed annually. Companies are also encouraged to adopt **Privacy by Design principles (PbD)** and to certify their data processing by a supervisory authority, possibly in cooperation with accredited third party auditors.
38%	42%	**Consent** must be given by a data subject in a clear statement or via an affirmative action (i.e. ticking a consent box when visiting a website) in cases when explicit consent would be required.
35%	43%	Serious **data breaches must be notified** to both the Data Protection Agency and data subjects. Supervisory authorities will maintain a public register of the types of breach notified. Notification must be given without undue delay.
21%	23%	A **data protection officer** (DPO) must be appointed by public authorities and businesses if data of more than 5000 data subjects is processed in any consecutive 12-month period. A DPO will also have to be appointed if (i) special categories of data, (ii) location data, (iii) data relating to children, or (iv) employee data in large scale filing systems are processed.

There are no fundamental differences between the perceptions of organisations making money by selling, analysing, or storing data (see table 47) and the rest of organisations, even though the

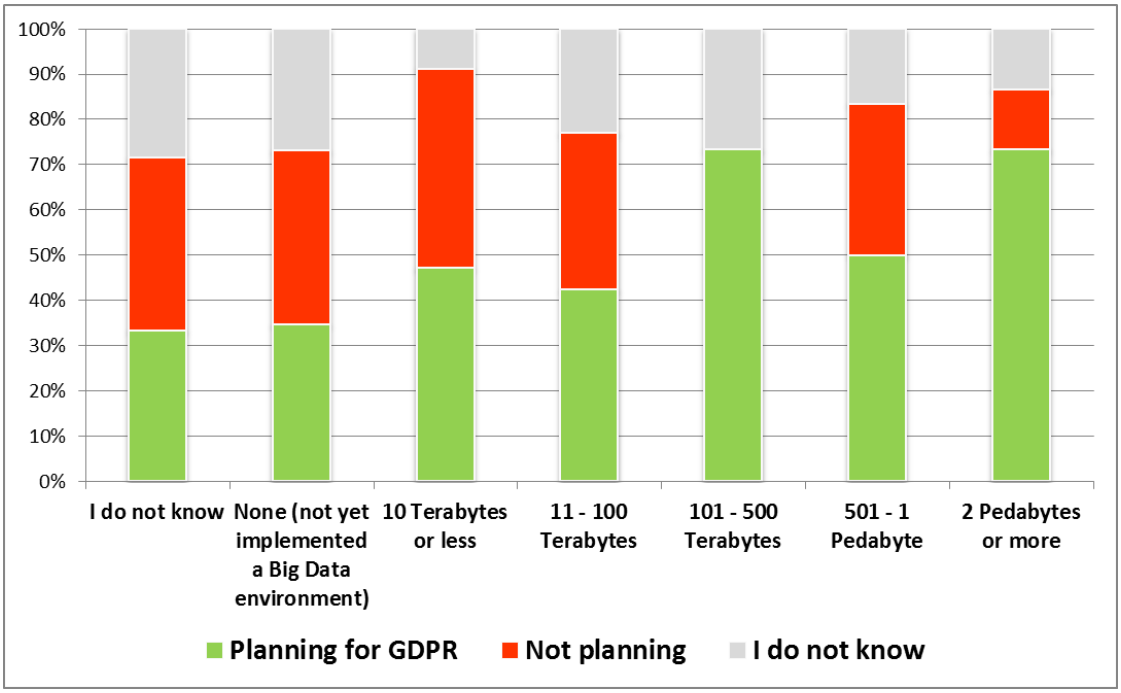
former ones seem to find slightly more problematic the provisions on data portability and explicit consent than the others.

Table 47. Special types of organisations: Percentages

	For Profit			Nonprofit	
	No	***Yes***		***No***	***Yes***
My organisation promotes or sells its products or services on Internet.	20 %	80%	My organisation promotes its services through a website.	7%	93%
My organisation uses monitoring devices to track customers or other people (e.g. web cookies, RFID, smart CCTV).	50 %	50%	My organisation uses monitoring devices to track users or other people (e.g. web cookies, RFID, smart CCTV).	32%	68%
My organisation generates income by storing data for other organisations.	69 %	31%	My organisation organises fund-raising campaigns on Internet.	59%	41%
My organisation generates income by selling data.	79 %	21%			
My organisation generates income by analysing data.	50 %	50%			
My organisation is a ISP, hosting or cloud provider.	80 %	20%			
My organisation is in the online advertising business.	85 %	15%			
My organisation does not do any of the above.	87 %	13%	My organisation does not do any of the above.	91%	9%
N	***133***			***74***	

With regards to the way organisations cope with regulatory uncertainty, in total only 39% of respondents said their organisations had already started planning for the new Regulation (n = 166). As showed in figure 27, organisations which process a lot of data, in the order of hundreds of terabytes or petabytes of data, are more proactive than other companies and have already started preparing for the new regulation and introduced measures to anticipate the envisioned regulatory change.

Figure 27. Organisations already planning for the GDPR by volume of data processed



6.7 Conclusions

This chapter has dealt with the analysis of the data and has tried to address the research questions. Propositions and hypotheses, identified in Chapter Four, are tested in Chapter Six by means of path analysis and other test for assessing indirect effects. Results are reported in section 6.5 and confirm almost the totality of propositions but two; no support was found for the hypothesis which suggests that the data protection regulatory regime would prevent indiscriminate data accumulation. Nonetheless, the presence of a reliable data protection regulatory regime contributes to foster the organisational privacy culture, increase compliance with data protection principles and ensure the respect of data subjects' rights. Analytically sophisticated organisations also tend to invest more in fostering their internal privacy culture especially when they plan to invest in target analytics. The conflict between data protection and big data emerge around the topic of data accumulation. The more organisations reuse and merge different data streams, the more challenging it becomes for them to comply with data protection principles and to foster the organisational privacy culture. To fully understand the implications of these results, the next chapter focuses on the implications of these results for practice, policy making and future academic studies.

CHAPTER SEVEN

Discussion of results: Implications, limitations and future research

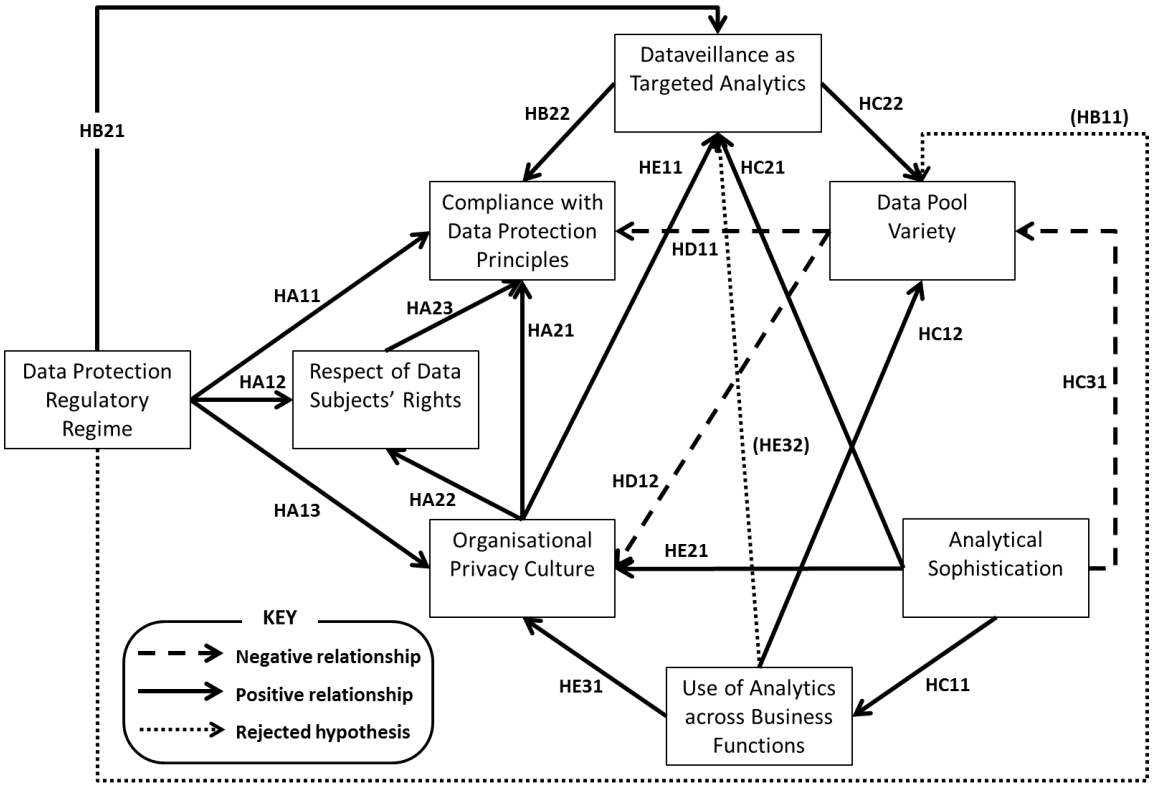
7.1 Introduction

Within this chapter, results presented in Chapter Six will be discussed, their implications explored, and their contribution to practice and academic knowledge presented. Limitations of this study and directions for future research are also included in this chapter.

7.2 Discussion of results

This study has tried to address the issue of the relationship between data protection and big data by looking at the experience of professionals working within organisations operating in Europe. Europe has been chosen as the context for this study because of its data protection legal regime, established by Directive 95/46/EC and corresponding national laws. As discussed in section 3.3, the European legal regime features basic data protection principles similar to the ones accepted in other contexts, such as in the US or in OECD countries. These principles set the rules for the lawful processing of personal data and the respect of data subjects' rights. The correct implementation of these principles within organisations is expected to improve data handling procedures and to allow data subjects' to exercise their rights to access, amend, or erase their data. This study has explored the relationship between the enactment of data protection laws and the level of compliance with data protection principles; the topic of the relationship between the degree of analytical sophistication an organisation has achieved and organisational data privacy decisions has also been investigated. The results of this investigation are discussed within this section and summarised in figure 28.

Figure 28. Summary of results: Hypotheses confirmed in the path analysis model



According to the results of this study, a strong and reliable data protection regulatory regime contributes to improve information management procedures within organisations through the enactment of data protection principles (HA11), and the respect of data subjects' rights (HA12). The regulatory regime contributes also to raise data protection awareness and to transform information privacy into a business priority, part of the organisational culture (HA13). The organisational privacy culture plays a fundamental role in transforming abstract data protection principles into practices and procedures within the corporate information management environment. The organisational privacy culture increases the likelihood of implementing measures which enable the organisation to enact data protection principles (HA21) and to respect data subjects' rights (HA22).

With regard to big data analytics, this study finds that targeted analytics is compatible with the current European data protection regime (HB21). To comply with the data protection regulatory regime, organisations know that they need to address information privacy issues at the time of using targeted analytics (HE11). Namely, organisations need to think how to comply with data

controllers' obligations at the time of designing their targeting strategies (HB22). This study found no evidence that data protection laws prevent data usage and accumulation (HB11). It also offers some insights about the relationship between big data analytics and big data. As expected analytically sophisticated organisations are those which use analytics (HC11) and targeted analytics (HC21) more extensively. This study also confirms results of previous qualitative studies (Davenport, Harris et al. 2010), which suggests that analytically sophisticated organisations understand the importance of respecting data privacy (HE21). The more they use analytics, the more they understand the need to establish information privacy as an organisational value (HE31). Accordingly, these organisations adopt a clear data gathering and handling strategy, tend to adopt principles such as data minimisation (HC31) to reduce the risks of data mismanagements, which help them reduce the problem of dealing with low quality data. Organisations which compete on analytics are clear about the purpose for which they collect the data and tend to process only the data they need. The problem is that the more organisations rely on analytics, the more they collect and merge data, the more difficult it becomes for them to comply with data protection principles (HD11) and to establish an internal privacy culture (HD12). And, as expected, the more organisations rely on analytics to achieve various business objectives (HC12), or on targeted analytics to profile customers (HC22), the more likely it will be that they collect a large variety of data. For this reason, this study claims that, although big data analytics and the current privacy regime are compatible, big data in its current interpretation of high volume, high velocity data in a variety of formats, is not well suited to accommodate data protection principles such as purpose limitations or limited data retention.

7.3 Implications

7.3.1 Implications for privacy studies

There is a growing need to understand how to manage customer privacy expectations in order to treat customers fairly and avoid discontent (Culnan 1993, Milne and Rohm 2000, Phelps, Nowak et al. 2000, Dolnicar and Jordaan 2007, Paine, Reips et al. 2007, Milne and Bahl 2010).

Organisations worry about the risk that advanced profiling procedures can become so intrusive to generate negative customers' reactions, boycotts, and massive withdraw of information (Lace 2005). Since it has been demonstrated that the perceived fairness of corporate information practices decreases customers' privacy concerns (Culnan and Bies 2003), retailers unable to credibly signal their trustworthiness in handling consumer information, might be less able to persuade consumers to share sensitive information necessary in the development of online commerce (Tang, Hu et al. 2008).

Companies have begun to recognise the importance of taking into account customers' reactions at the time of applying sophisticated analytics. Privacy and security scandals can ruin the reputation of organisations and their market value (Mulligan and Perzanowski 2007). Facets of privacy practices such as data retention, access to collected data, and scope of use affect users' willingness to allow the collection of behavioural data (Leon, Ur et al. 2013). Permissions display features and privacy options can play an important role in app-selection decisions (Kelley, Cranor et al. 2013). Investing in privacy awareness campaigns and information security training goes hand in hand with the use of targeted analytics. Technical solutions, such as privacy-preserving profiling, are also available to ensure compliance with data protection norms (Vaidya and Atluri 2007).

This study has investigated the corporate data privacy environment by paying attention to the organisational privacy culture and to the privacy regulatory regime, as done in previous studies (Smith, Dinev et al. 2011), while also introducing new concepts borrowed from business studies and surveillance studies such as Analytical Sophistication and Targeted Analytics interpreted as a form of Dataveillance. New insights on the relationship between data usage and data privacy have emerged from the interaction between these different streams of research. The contribution of surveillance studies has also helped the researcher identify potential limitations of the current privacy regime.

An area of concern when reflecting on the relationship between data protection and big data is data accumulation. According to the results of this study, it seems to be fairly problematic for organisations which process a large variety of data to comply with data protection principles (HD11) and establish an internal privacy culture (HD12). Analytical competitors seem to be ambivalent with respect to data accumulation. On one hand, as suggested in previous studies (Davenport, Barth et al. 2012, Davenport and Dyché 2013), they understand the importance of safeguarding information privacy (HE21). Yet, on the other hand, the demand for data fusion and accumulation grows with the use of analytics. Thus, analytically sophisticated organisations are more likely to gather, store and analyse a huge quantity and variety of data to solve a large variety of business problems within the organisations. Data accumulation trends create frictions with data protection principles, especially with the purpose limitation principle (Cate, Cullen et al. 2013), and lead to a decrease in the chances of developing a strong organisational privacy culture. A lot of attention has been paid in privacy studies in assessing privacy policies and consent-based procedures. Future studies should address the specific trade-offs between other data protection principles or legal provisions and specific technological developments.

Since compliance pressures may increase the demand for information security products and services (Khansa and Liginlal 2007), there is a need to explore those institutional and organisational elements which contribute to stimulate technological, procedural, and organisational innovations in the area of data privacy and security. There is also a tendency towards the development of additional technologies to limit, coordinate or manage existing technologies, which might be interestingly interpreted as a form of technological super-fix (Wynne 1975). As it has been pointed out by some commentators, "[t]here is substantial appeal in the idea of a technological solution to a problem that technology itself seems to have created, in part because such technologies are self-enforcing and appear to reduce the need for regulatory interventions" (Samuelson 2000: p. 1167). Empirical studies on the development of privacy invasive or enhancing technologies (PETs), despite their relevancy and potential practical implications (Cranor 1999), are still very limited. Specific studies are needed to assess both the

design of information privacy devices, their impact on potential users (Bélanger and Crossler 2011), and the advantages of adopting a technological solution in comparison with a non-technological alternative. Initiatives such as the Platform for Privacy Preferences (P3P) can help users and organisations communicate by means of automated mechanism to check privacy policies against users' preferences (Baumer, Poindexter et al. 2004). New techniques which enable the alignment between privacy policies with system requirements are constantly developed (Antón, Earp et al. 2003). Nonetheless, the automation of these procedures poses perils as they require the translation of legal concepts into machine-readable commands (Bamberger 2010). Through the adoption of a multidisciplinary approach, future privacy studies may try to address the limits and opportunities of tackling information privacy problems through the adoption of solutions which feature both technical and legal elements.

Finally, most privacy studies rely on quantitative methods. The present study is no exception. More qualitative research is probably needed in order to understand how different stakeholders, such as law makers, security experts, software developers, analysts, compliance officers, C-level executives, frame issues related to information privacy and data usage, and to identify potential spaces of dialogue.

7.3.2 Implications for the information security literature

Finally, as presented in section 6.6.1, this study also contributes to the information security investment literature (Chai, Kim et al. 2011, Lee, Kauffman et al. 2011), by exploring the rationale behind information security investments. According to the preliminary results reported in this study, organisations can follow either a more reactive or a more proactive approach toward investing in information security depending on whether they value data security as a value-added attribute or not. Organisations which operate in sectors characterised by strict information security requirements, seem to have an incentive to invest in data privacy far beyond compliance because their managers perceive to face serious reputational risks. Managers in other organisations tend to assume a more reactive posture and invest in information security only if

something, such as a serious data breach which changes perceptions, happens (Volpentesta, Ammirato et al. 2011). The risk of being sanctioned by regulators, or sued by clients, may also motivate investments. These considerations have important policy implications and suggest that competitive pressure helps increase information security procedures more than regulatory sanctions, even though sanctions are necessary to increase information security levels within all kind of organisations (Nettleton and Turner 2008).

As showed in this study, big data analytics is used by organisations to tackle information security issues. Information systems research in the areas of big data infrastructure, analytics and decision support systems, can be improved through more interdisciplinary collaboration (Goes 2014) with scholars doing research on information security matters. This study also shows that information security and data privacy are strictly intertwined phenomena and that a more strict collaboration between the privacy and security academic communities would produce important advances in both areas of inquiry.

7.3.3 Implications for surveillance studies

Surveillance has been interpreted as an everyday, ambiguous experience, and a new form of modern governance characterised by the reproduction of social divisions through the manipulation, decontextualization and classification of information about individuals (Lyon 2002). The intrinsic 'ambiguity' of any manifestation of surveillance is considered to derive from the contested nature of the functions surveillance fulfils. In a world of strangers and "disappearing bodies from integrative social relationships" (p. 243) socio-technical surveillance systems deliver abstract actuarial tokens of trust—or supposed 'justified' suspicion—which shape social relationships. The politics behind these classifications, maybe because they are run by 'neutral' scientists or automated procedures, is always hidden and often unknown. And, as there is no single bureaucracy or institution, no overreaching Big Brother to blame, and because of its unstable, ubiquitous and amorphous nature, the 'surveillant assemblage' seems to pose big challenges to a satisfactory comprehension of it (Haggerty and Ericson 2000).

This approach offers a theoretical framework for understanding both the roots and the long-term, ethical and societal consequences of implementing overwhelming surveillance measures. On one side, surveillance comes from the capitalist drive for greater profit, the functioning of rational bureaucracies and the disciplinary power over the self, exercised by modern institutions. On the other, surveillance may lead to cumulative disadvantage, such as inter-generation income inequality or uneven distribution of life chances as an effect of stigmatization (Gandy 2010). The importance of the implications coming alongside the capillary reliance on surveillance practices, as an ordinary mode for organizing social relationships for the construction of our future societies, motivates our attempt to frame the theme of data management as a problem of surveillance, rather than just as a problem of privacy.

But, if theories about privacy tend to assume an individualistic perspective, surveillance theories go far beyond the organisational level to ask questions about social order and social reproduction. In order to reconcile this two opposite viewpoints, this study has referred to a particular manifestation of informational surveillance which underlines both the methods used for collecting data, which often rely on several surveillance procedures and technologies, and the outcome of this investigation, that is information organised in databases. The practice of monitoring population through the recoding of digital data is called *dataveillance* (Clarke 1988).

According to the results of this study (HB21), and as pointed out by surveillance scholars, the current privacy regime seems to support the expansion of dataveillance procedures such as profiling (Gilliom 2011, Degli Esposti 2014). By becoming part of the corporate narrative, the notion of privacy represents a space of encounter and confrontation between those in favour and those against the proliferation of mass surveillance (Coll 2014).

By drawing insights from sociology and social psychology, surveillance studies can strongly contribute to move the privacy debate in more fruitful directions. By paying attention to group dynamics, rather than individual preferences, surveillance scholars can shed light on the so-called

"privacy-paradox" and show how particular groups, such as young people, who heavily rely on digital technologies, understand privacy and react to digital surveillance.

The consequences of dataveillance for society may be beneficial in some respects, but also detrimental in others. Businesses are expected to create value, by means of dataveillance, through higher sales or personalisation of services. But the most important aspect related to the use of this concept comes from being a self-critical idea that does not take for granted the necessity and inevitability of any type of surveillance practice. More empirical studies on all aspects of dataveillance are still needed. New solutions must be identified. Otherwise, the most meaningful choice made by Internet users to safeguard privacy would remain to falsify information when personally identifiable information (PII) is requested by Web sites (Baumer, Poindexter et al. 2004).

Finally, there is a need to shift to an integrated political economic analysis of surveillance capable of exposing complementarities and synergies between state and corporate priorities, intimately linked with developments in the natural and applied sciences (Ball and Murakami Wood 2013). Another aspect completely overlooked by those perspectives which understand privacy as a mere information disclosure exercise is that asymmetries of information typically heighten power imbalances and put individuals at a distinct disadvantage (Cavoukian and Kruger 2014). The study of privacy from a surveillance studies perspective can help uncover the social structures and the power dynamics which are emerging in the context of the digital interaction between users and organisations. This study has tried to make a contribution in this direction by exploring the interplay between the current privacy regime and corporate priorities.

7.3.4 Implications for practice

This study confirms that investing in building an organisation's information privacy culture is a necessary condition to become a successful, analytically sophisticated organisation. According to the results of this study (HC31), analytically sophisticated organisations try to apply data minimisation solutions. Data discovery and data quality are important big data challenges.

Identifying high-quality data from the vast collections of data that are out there on the Web is not straightforward (Zicari 2013). So, while there are forces intrinsic to the design of the big data IT infrastructure which push toward the reduction of data through the elimination of irrelevant and outdated data, in line with data protection principles, the rationale behind data analysis lead organisations to continuously look for new data sources and the integration and retention of this information in search of novel applications. For this reason, when analytics enters into the picture, the amount and variety of data processed by the organisation increase dramatically. These considerations have important implications for those who design the information management system. The importance of building privacy into the system's architecture becomes more evident after looking at the results of this study (propositions D, C, and E).

The more organisations rely on analytics, the more they demand data; and the more they demand data, the more difficult it becomes to establish an internal privacy culture. These considerations have important implications also for privacy professionals who are trying to make a case in favour of information privacy within their organisations. By framing the issue of data privacy in terms of data quality, privacy professionals working in analytically sophisticated organisations can persuade their colleagues working in the marketing and information systems departments of the value of applying basic data protection principles, such as data minimisation.

To enable organisations to develop an internal privacy culture, economic and human resources must be allocated to pursue this objective. As explained in section 6.6.3, privacy issues are commonly managed by a designated department within the organisation. Increasingly it is becoming a common practice to appoint a Chief Privacy Officer (CPO) leading the privacy team. The privacy team needs to be able to talk the language of information security practitioners, besides dealing with privacy laws and compliance issues. It also has to engage with employees to educate them to handle data appropriately.

Finally, big data and analytics, despite being often proposed as a panacea, pose important challenges and leave unspoken questions, such as: how good is the data? How broad is the

coverage? How fine is the sampling resolution? How timely are the readings? How can we cope with uncertainty, imprecision, missing values, misstatements or untruths? (Zicari 2013). Topics such as information security and data privacy play an increasingly important role in the way big data projects are designed and handled. Some organisations are already moving toward the integration of the privacy and information security functions. As CIOs are being converted into Chief Innovation Officers (Goes 2014), they should foster the dialogue between privacy and security professionals to help identify new ways to understand privacy and to move on from the old discussion on 'data control' to 'accountable data use' (Weitzner 2006).

7.3.5 Implications for policy makers

According to the results of this study (HA11), the European data protection regime represents a point of reference not only for privacy advocates, but also for private firms which have made a number of internal procedural changes to improve their data privacy standards and comply with the law (Shaffer 1999, Shaffer 2000). So, the governance of privacy, when effectively implemented and strongly enforced (HA12), may help people exercise their data protection and human rights (Bennett 2011). Organisations which implement procedures to ensure the respect of data subjects' rights are also more likely to fully comply with data controllers' obligations (HA23).

As governments become more involved in the corporate management of information privacy, and according to proposition A, the internal management of such issues seems indeed to tighten (Milberg, Smith et al. 2000). Furthermore, this study confirms that "[t]he notion that privacy must be sacrificed for innovation is a false dichotomy, consisting of unnecessary trade-offs. In fact, the opposite is true: privacy drives innovation. It forces innovators to think creatively to find solutions that will serve multiple functionalities" (Cavoukian, Stewart et al. 2014: p. 16). This study found no evidence to support the idea that data protection laws hinder innovation or create disruptions to data flows (HB21 and HB11). As already demonstrated in previous studies, privacy laws have not stopped international data flows as originally thought (Samiee 1984). These findings confirm

that European data protection laws do not create major disruptions or concerns to corporations (Kane and Ricks 1988). The current data protection regime does not create disruptions to the proliferation of big data analytics solutions, especially to the application of targeted analytics (HB21). The regulatory regime forces organisations to address data privacy issues, especially when they want to make use of targeted analytics (HE11).

Although big data analytics seems to be compatible with the current privacy regime, the rhetoric around big data, characterised by indiscriminate data accumulation, is virtually incompatible with core European data protection ideas. As pointed out by a few commentators (Cate, Cullen et al. 2013, Mayer-Schonberger and Cukier 2013), organisations want to retain data indefinitely and want to be free to merge all data they possess in search of answers and new lucrative applications because "the advent of Big data and new analytical tools has shown us that many valuable and innovative uses of data are not known at the time of collection" (Cate, Cullen et al. 2013: p. 7). Accordingly, some experts consider inappropriate and unproductive to maintain the purpose limitation principle in future data protection legislations (Cate, Cullen et al. 2013, Mayer-Schonberger and Cukier 2013). The study makes a case, and offers evidence, against this idea; thus it supports the vision of the regulator (Art29 2014). On one hand, the real value of big data comes from big data analytics, which is compatible with the current data protection regime. On the other hand, organisations need to organise and understand what data they collect, assess the quality of these data and protect them from unauthorised access. Ideas such as data minimisation help organisations manage their data effectively (HC31). This study contributes to the debate on the necessity to maintain the purpose specification and limitation principles in the proposed General Data Protection Regulation, by emphasising the beneficial outcomes that the existence of data minimization ideas in future legislation can have on data quality and data security.

Despite the rhetoric about big data, organisations need to take important decisions about the data they store; decisions about data access and confidentiality are particularly important these days. The frequency of privacy disasters might increase with the growth and accumulation of digital data; to cope with these challenges organisations need to find the right mix of measures

and the adequate legal solutions to reduce data privacy risks (Chan, Culnan et al. 2005). Infamous data breaches, such as the Ashley Madison dating site case, should serve as an example of the kind of reputational risks a data breach can bring to organisations. In the Ashley Madison case, for example, several individuals were using their job email account to access the service. A clear and strong data protection regulatory regime would help all kind of organisations increase their information security and privacy standards.

Further evidences of the need to establish privacy legislation to set minimum information privacy standards has been provided in section 6.6.1, where drivers of information security investments were discussed. According to the results of this study, organisations tend to adopt either a proactive or a reactive posture toward information security. Proactive organisations work in highly regulated sectors where information security is perceived as a quality dimension of the products or services they offer. Reactive organisations, in contrast, invest in information security after suffering a major data breach, or when they face the risk of being fined by the regulator or the risk of having to pay high litigation costs. Privacy laws should be designed in such a way to persuade organisations to adopt a proactive attitude toward information security. The protection of data privacy and information security should not be perceived as mere compliance issues; it should rather become part of quality assurance procedures. Otherwise, the introduction of specific privacy requirements in public procurement could help the market for Privacy Enhancing Technologies (PETs). Provisions regarding Privacy Impact Assessment, PETs, Privacy by Design, and similar ideas included in the draft Regulation will probably help organisations begin to make use of these measurements. According to the results of this study presented in section 6.6.3, a limited number of organisations rely on these solutions nowadays.

Finally, according to the results presented in section 6.6.4, the right to erasure and the right to data portability raise concerns within organisations. Most professionals are also unfamiliar with legal instruments such as Binding Corporate Rules and consider their adoption problematic. Organisations which process considerable amounts of data have also already started preparing for the new Data Protection Regulation, whose final text has been published on the 4th of may

2016 on the *Official Journal of the European Union, L 119* . These results offer further evidence on the impact of privacy legislation on organisational decisions.

Data breach notification laws could also help increase minimum information security standards across sectors. As discussed in section 6.6.2, by recognising data breaches as a privacy problem (Culnan and Williams 2009), private and public entities could speed up the process of abandoning a reactive approach toward information security investment decisions to move on to a more proactive approach which understands privacy and security as quality attributes of any digital service or product.

7.4 Methodological implications and limitations of this study

This study has tried to explore the relationship between big data and data protection by looking through the eyes of privacy professionals and IT practitioners. The empirical model presented in section 7.2 highlights potential trade-offs and synergies between the way data protection law and big data technologies are understood by survey respondents.

From a methodological point of view, this study responds to the call for specifying boundary conditions for sample-based general knowledge claims (Seddon and Scheepers 2012), to the need to address considerations related to measurement model specifications (MacKenzie, Podsakoff et al. 2005), and to the requirement of adopting statistical techniques whose assumptions match the nature of the data analysed (Wasserman 2006). By relying on a quantitative research design, this study has also tried to offer a systematic view of the relationship between constructs coming from different research traditions. It also complements and enriches the knowledge produced in previous qualitative studies in the field of surveillance studies (Ball 2010), privacy studies (Bamberger and Mulligan 2011), and analytical competitor studies (Davenport, Barth et al. 2012).

The validity of these results is based on the implicit assumption that survey participants were both knowledgeable about their organisations'' data management and analysis procedures, and willing to provide accurate and truthful answers about these issues. Although the original

objective was to investigate perceptions of professionals based in different European countries, the sample finally obtained features a large proportion of professionals working in British organisations. Although study participants came from organisations of different sizes and from various sectors, the researcher acknowledges the limits to the generalizability of results produced by the data collection strategy adopted, as already discussed in section 6.2, and the importance of understanding context characteristics (Davison and Martinsons 2015).

Some further limitations of this study need also to be addressed. First, as the study was based on people's perceptions, rather than on factual information related to each respondent's organisation, and the survey was completely anonymous, the researcher was unable to contrast information provided by respondents against official records. On the other hand, a survey represented a good opportunity to obtain information on a wide range of issues related to both legal compliance and information systems characteristics. Future research could overcome these limitations by means of a mixed-method research design which could include in-depth interviews, survey data and official documents.

Another limitation, typical of self-administered questionnaires, is that the researcher was unable to control for multiple interpretation of the questions included in the survey. As pointed out by Chris Tiernan, member of the ELITE group of the British Computer Society, the word 'infrastructure' in question Q21_2 "My organisation has a flexible, centralized IT infrastructure to work with data Vs. My organisation lacks a flexible, centralized IT infrastructure to access and work with data", part of the Analytical Sophistication scale, could have produced some confusion. According to Mr Tiernan, the information management system and the underlying IT infrastructure are completely different things in that the information management system could be fully integrated, whilst the underlying infrastructure consists of many different components, e.g. some on the premises and some in the cloud and a variety of technologies. What respondents think infrastructure means could have influenced their answers.

Study participants gave their feedback on the survey by using the space for comments positioned at the end of the electronic questionnaire. Some of these comments, here reported, highlight other potential shortcomings.

- “No discussion of 'risk' anywhere here – possibly the most important aspect of privacy management.”
- “Hard to answer, as some questions have multiple answers depending on the type of data.”

Some participants wanted also to share their views on what kind of questions should have been included in the survey or on how specific questions should have been formulated.

- “I would have expected more on 1) the attitudes & expectations of our customers/clients & 2) the potential conflict of interest (e.g. concerning deletion or correction of information) between persons who are subjects of data, & the formal record-keeping of professionals whose records (that are data) contain personal information about clients.”
- “Regarding the question on how problematic certain principles can be, the answers given relate to practical implementation, NOT to the need for (similar) principles.”

Both the ‘relevant media’ sampling strategy finally adopted, and the convenience sample obtained as a result, pose additional limitations to the generalizability of the results of this study. As explained in section 6.2.1 the researcher could not compute the response rate as she had no information on the sample frame. Respondents were also mostly based in the UK (76%; n = 195). British respondents were also more critical, on average, of their national data protection authority than respondents based in other countries. As reported in the Participants’ Study Report included in the Appendix, three quarters of UK-based survey respondents said that the Information Commissioner's Office (ICO) did not have the power, or the resources, to impose serous sanctions and that privacy law was enforced in an inconsistent and unreliable manner. Although these respondents declared that they were knowledgeable about data protection law, there is a risk that the sample contained an overrepresentation of highly critical, or unsatisfied,

professionals with a specific vision of how privacy law should be administered. The following comment written by another survey participant offers an example of the attitude toward the ICO of respondents.

- "Big issue is the ICO appear to pre-empt what is important which creates a self-fulfilling result. A big example of this was Cookies – the law was explicit consent and need to ensure understand each cookie consenting to yet the ICO early on said that enforcement would be low and only against severe breaches as people weren't bother about needing to consult on each cookie – now quote as saying few complaints as evidence of right in original thought – yet failed to realize that people just don't put complaints in because they feel ICO not interested as already said as much. The regulator also has insufficient ability to fine – amount low – compare with FCA and fines miniature and firms likely to see profit as clearly outstripping any penalty – very similar to old FSA fines which were changes so breach costs were substantial and a real risk to encourage compliance."

To overcome this limitation, this study should be replicated in other European countries. The researcher tried to obtain more responses from other European countries, such as Spain, without being successful.

Another unexpected problem generated by the composition of the sample was the over-representation of private firms which operate in the business-to-business market (For profit organisations: 68%; n = 99; Nonprofit organisations: 16%; n = 58). This issue was pointed out by another study participant, who wrote in his/her comment: "the direct impact on our business will be minimal as we manage largely B2B data. However, we work with many clients who have large consumer databases and a number who resell customer data; as a result, we are very concerned about some of the proposals of the Directive. In my experience, few of our customers are in any way ready to implement the proposed changes and most would find it extremely problematic to do so."

To address most of the problems just mentioned the researcher envisioned to run a follow-up to the study. For this reason, study participants were asked if they were interested to be contacted again by the researcher. 35% of respondents said they wanted to be contacted again either by email or phone (n = 159). Unfortunately the researcher had no time to contact these people and add this additional layer of information to the study.

With regard to specific methodological aspects, there was no possibility to validate the model by using a different sample from the one used to develop the instrument. The fact of not being able to use latent variables and to build a proper structural model because of the limited sample size has also created problems as the model finally presented does not account for measurement errors. The spurious variance attributable to the measurement method rather than to the constructs the measures are assumed to represent, a phenomenon known as common-method variance, could also have increased the correlation between variables. The use of path analysis and procedures for testing mediation effects with cross-sectional data also raises concerns on the possibility to have a final word on the directionality of relationships and the accuracy of parameter estimates (Maxwell and Cole 2007).

These limitations notwithstanding, this study provides some of empirical evidence of the interplay between regulatory requirements and organisational decisions in the area of big data analytics and data protection from the perspective of survey respondents. This study also responds to the call for studying privacy at organisational level and conducting more interdisciplinary privacy studies (Pavlou 2011); namely, in order to study organisational data privacy decisions, the privacy studies literature has been complemented with insights drawn from surveillance studies. This study is also a first attempt to develop scales to measure the level of compliance with data protection principles.

7.5 Conclusions

Data is becoming such a central resource for our economies in the 21st century to be compared with what steam power has been for the 18th century, electricity for the 19th century, and

hydrocarbons for the 20th century (IBM 2013). Setting a distinctive balance between the private sphere and the public order in the digital age is a difficult exercise (Westin 2003). Organisations and law makers struggle to strike a balance between data usage and data protection. The more internet-mediated interactions become widespread and increasingly sophisticated data gathering devices are developed, the more data become available on the market. New types of service companies emerge to exploit the potential of digital data. These companies, called generally data firms in the context of this study, make profits through the collection, assemblage, sale, and analysis of data. Data, however, refer quite often to identifiable persons who can suffer serious damages in case data get lost or disclosed to unauthorised third-parties. To safeguard people's privacy and protect data secrecy and integrity, several laws have been enacted across developed countries to force public and private institutions processing data, to comply with the so-called data protection principles. According to these norms, data-subjects, defined as the person the data refer to, must be informed about the collection and use of their personal data. Data-subjects also have the right to deny, or grant, consent to the collection and processing of their personal data as well as the right to access their personal data to be sure they are accurate. Finally, data-controllers/processors—i.e. the ones who are collecting and processing data—have the duty of secure information from theft or abuse and they are considered liable for any occurred security breaches.

From economics to business studies, and from law to public opinion pool research, many scholars have analysed why and under what circumstances people are concerned about their privacy. Very few studies have paid attention to the study of the corporate data privacy environment as well as to the interplay between privacy regulations and business data protection choices. There are evidences, however, suggesting that companies, which operate under a strict regulatory privacy regime, tend to implement more protective internal privacy-preserving measures (Milberg, Smith et al. 2000). Although it has been noticed that in case of emerging conflict of interest between the use of customers' data for making an extra sale employers and employees may make an exception to data protection rules (Ball 2010), there was a need to run a systematic study on the

relationship between dimensions of value creation, through data analysis and brokerage, and the level of data protection.

Thus, this study has represented an attempt to shed light on the relationship between characteristics of the privacy regulatory regime, the level of analytical sophistication achieved by an organisation, and the internal corporate privacy culture. Results are based on 195 usable surveys out of 442 returned surveys. Organisations represented in the sample were mostly based in the UK. Out of five overarching propositions, twenty hypotheses have been tested by means of path analysis and other tools to assess the effects of intervening variables. Additional analyses based on distribution-free statistics have also been performed in order to explore corporate information security investment decisions, relationships between privacy and security safeguards, and organisational reaction to provisions contained in the proposed European General Data Protection Regulation.

According to the results of this study, a strong and reliable data protection regulatory regime contributes to raise information privacy standards within organisations, through the adoption of basic data protection principles. The more organisations make an effort to build an internal privacy culture, the more likely it becomes that legal provisions are translated into corporate practices. Organisations which rely on analytics to achieve business objectives, and in particular rely on targeted analytics, are more likely to develop an internal privacy culture. Although analytically sophisticated organisations tend to apply data minimization rules and other privacy-preserving methods, they face the challenges of balancing the demand for data fusion and collection with data protection considerations. The more organisations apply analytics to achieve different business objectives, the more it increases the amount and variety of data gathered and processed by organisations.

This research contributes to the information privacy literature in two ways: it investigates the corporate data privacy environment, a topic which has received limited attention; it also draws insights from surveillance studies to explore the European regulatory privacy regime and improve

our understanding of the relationship between the usage and the protection of personal data. Finally, the study's findings are relevant to both practitioners and policy makers, as they offer evidences of the positive relationship between privacy regulation and business competitiveness, and of the positive effects of privacy regulation on overall organisational information handling procedures.

Statistical Appendix

SA.1 Statistical approach: Nonparametric methods

Parametric statistics is commonly chosen to run multivariate analysis in business studies and social sciences (Wasserman 2004). The problem with parametric statistics is that it makes restrictive assumptions about the shape of a population distribution when performing the hypothesis test. The reliance on normal theory implies that four basic assumptions must be met. Considerations related to each of these assumptions are made and reported in table 48.

Table 48. Parametric statistics' basic assumptions

Parametric statistics assumptions	*Problems with meeting these assumptions in this study*
First Assumption Variables must be measured on an *interval or ratio scale*, which means that the absolute distances among levels, must be known.	Within this study, using variables measured on an ordinal scale, like in the case of a Likert scale, would not be considered optimal as, though there is a clear ordering of the levels, the absolute distances among levels are unknown (Agresti 2010). The 7-point bipolar scales used in this study suffer from the same problem.
Second Assumption *Participants should be randomly selected*: possible participants should have an equal likelihood of being selected for participation in the study and they should represent a random sample of the targeted population.	Random selection of participants is a condition which does not hold in the case of a convenience sample (Linebach, Tesch et al. 2014).
Third Assumption *Responses should be independent*: data must be orthogonal, which means that "one variable has no impact on another variable; one participant's response has no impact on another participant's response" (Linebach, Tesch et al. 2014): 22.	As participants were filling in the questionnaire electronically from different locations and at different time there is no reason to believe that their responses were influenced by other study participants or by any third common intervening factor.
Fourth Assumption The variance of two or more groups or samples, also known as *homoscedasticity or* homogeneity of variance*,* should be equivalent (Linebach, Tesch et al. 2014).	This issue is addressed in section 6.4.In addition tests on the equality of standard deviations (variances) can be performed.

Another problem with parametric statistics is that having multivariate normally distributed data obtained from a survey is always a challenge. Omnibus tests for multivariate normality have been

applied to study the multivariate distribution of the all the main variables (Mardia 1970, Henze and Zirkler 1990, Doornik and Hansen 2008). These tests, performed in Stata 8 (Stata 2015), provided evidence that the variables used to measure the constructs presented in Chapter Five were not-normally distributed. The complete list of variables and constructs with descriptive statistics is reported in the statistical Appendix.

Researchers tend to rely on the *Central Limit Theorem* and the *Law of Large Numbers* to cope with the problem of not having perfectly normally distributed data because, according to these theorems, the sampling distribution of the mean approaches a normal distribution as the sample size increases. It seems that "sample sizes above 30 should generally be large enough for the Central Limit Theorem to be used" (Hubert and Wainer 2012): 176. In brief, as the sample size increases, the estimator of the true population mean converges to the true mean at the limit as the size of the sample goes to infinity. This is seen most directly "in the variance of the sampling distribution for the sample mean becomes smaller as the sample size gets larger" (Hubert and Wainer 2012): 177. As a result, averages are both less variable and more normal in distribution than individual observations. Furthermore, averages based on larger sample sizes will show less variability than those based on smaller sample sizes.

Thus, the approximate confidence interval statement remains valid even when the underlying distribution is not normal. Such a result is the basis for many claims of robustness stating that the procedure remains valid even if the assumptions under which it was derived may not be true, as long as the sample size is reasonably large (Hubert and Wainer 2012). However, since data are not normally distributed and, thus, normal theory assumptions have been only partially met, it would be advisable to abandon the parametric approach and adopt nonparametric statistics (Hollander, Wolfe et al. 2014). In addition, as sample statistics are used to estimate population parameters, it remains more appropriate to extract a representative random sample from a population in order to use parametric statistics, something it was not possible to achieve in this study.

Nonparametric statistics, also called *distribution-free statistics*, provides a viable alternative to parametric statistics (Klein 2011) as they are better suited for testing hypotheses while dealing with ordinal data. In fact, while the parametric procedures require the magnitude of the observations, the advantage of using nonparametric procedures is that they often require just the ranks of the observations (Hollander, Wolfe et al. 2014). The problem with nonparametric tests is that they are often not as efficient – or sharp – as parametric statistics. However, nonparametric procedures are only slightly less efficient than their normal theory competitors when the underlying populations are normal, but "they can be mildly or wildly more efficient than these competitors when the underlying populations are not normal" (Hollander, Wolfe et al. 2014: p. 1).

As survey data collected as part of this project are visibly not-normally distributed in conjunction with the fact of having adopted a convenience sampling procedure, these concerns led the researcher to conclude that in the context of this study it would be more appropriate to use nonparametric procedures. Table 49 offers a comparison and a summary of parametric and nonparametric statistics.

Table 49. Nonparametric vs. Parametric Statistics

Parametric statistics	Vs.	Nonparametric statistics
Preconditions		
Dependent variable: Variables at least interval scale (i.e. interval or ratio)		Dependent variable: nominal or ordinal Replace data with ranks
Assumptions		
1) Randomly sampled data 2) Independent sampling 3) At least interval data 4) Homogeneity of variance 5) Normally distributed data 6) Need to check also for outliers and nonlinear association between variables		Distribution-free: data are not assumed to have any characteristic structure or follow any predetermined distribution.
Advantages		
1) Parametric statistics are used to make inferences about population statistics 2) They offer higher power for your tests: it would be more likely to reject the null hypothesis when it is false 3) Possibility of applying the Central Limit Theorem with large sample		1) Exact p-value for small sample size and exact confidence intervals 2) Since it works on ranks, it does not require numerical data but can be performed on ordinal data 3) It provides simple tests for complicated hypotheses 4) More robust statistical procedure 5) It protects against the influence of outliers
Disadvantages		
Need to respect assumptions		Tests are considered less powerful, though it depends on the underlying distribution and on the test performed
Equivalent measures of association		
Chi-square test		McNemar's test Fisher's exact test
Pearson's product moment correlation coefficient (-1 < R < 1) of linear association.		Spearman's rank correlation coefficient (rho) Kendall's tau
Linear regression		Non-parametric regression

SA.2 Complete list of survey items measuring each construct with descriptive statistics

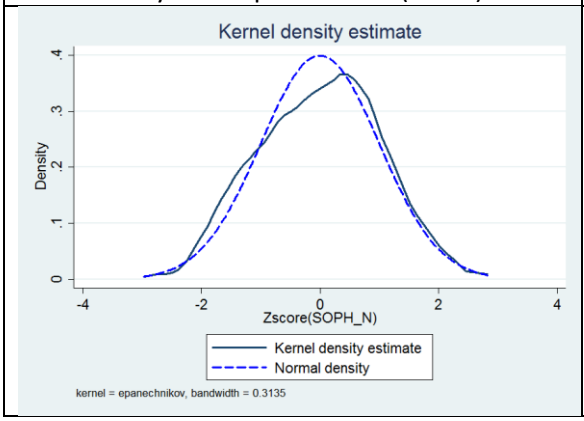
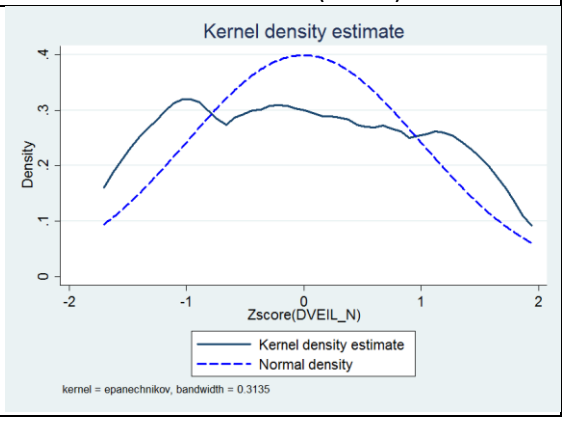
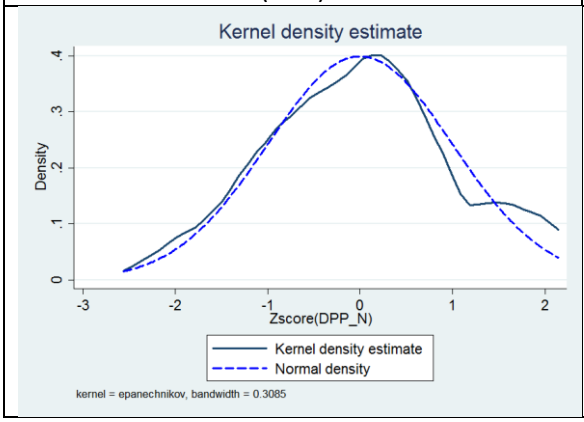
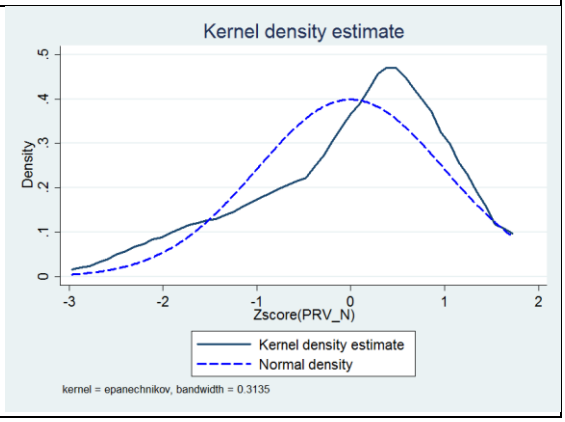
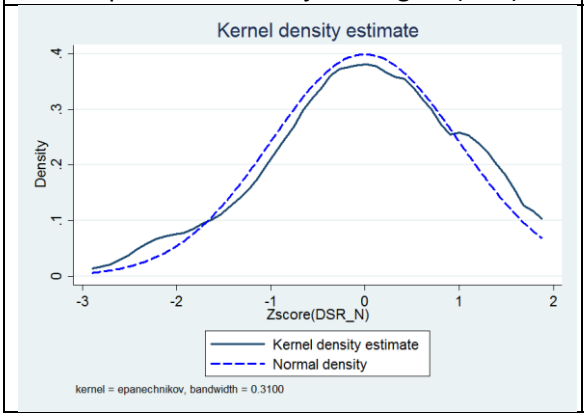
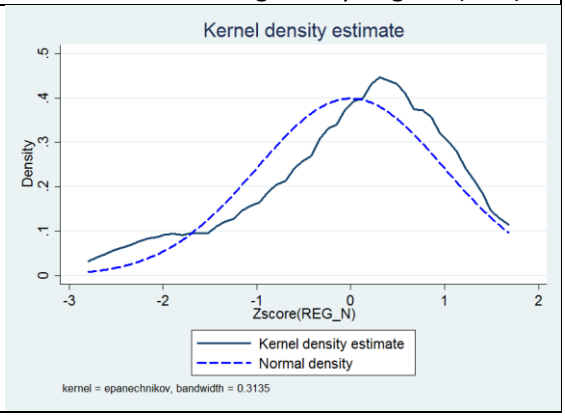
Table 50. List of constructs and corresponding variables with descriptive statistics

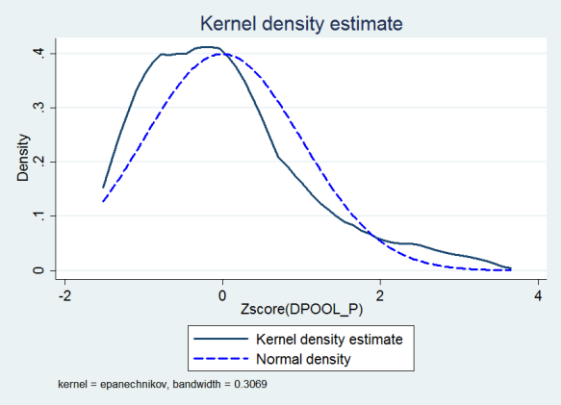
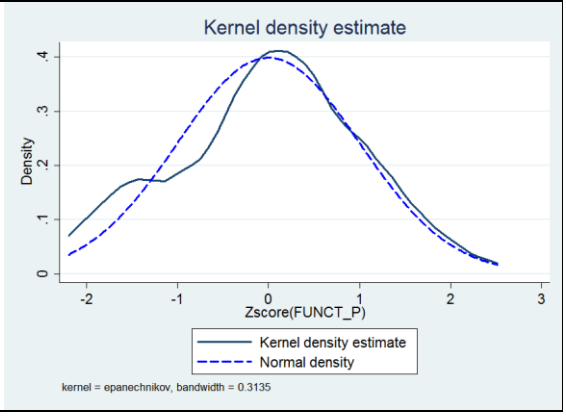
Variable	Survey item	Obs.	Mean	Std. Dev.	Min	Max
	Analytical Sophistication (SOPH)					
Q21_1	Within my organisation, data are accurate, stored in compatible formats and easily accessible.	257	3	3.25	1.62	7
Q21_2	My organisation has a flexible, centralized IT infrastructure to work with data.	258	3	3.23	1.67	7
Q21_3	Employees are encouraged to rely on data analytics.	255	4	3.84	1.73	7
Q21_4	We have analysts able to mine data and get useful insights.	253	4	3.63	1.96	7
Q21_5	Data analytics represents a distinctive, competitive capability of my organisation.	256	0	0.38	1.93	7
Q21_6	Digital data represents a core asset, key to our business model.	258	3	3.00	1.73	7
	Dataveillance as targeted analytics (DVEIL)	Obs.	Mean	Std. Dev.	Min	Max
Q29_1	My organisation collects data to monitor individuals' activities.	174	4	3.98	2.27	7
Q29_2	We analyse personal data to foresee and influence people's behaviour.	175	4	4.30	2.26	7
Q29_3	Profiling is used to target valuable users or personalise offers.	173	5	4.53	2.26	7
	Data Pool Variety (DPOOL)	Obs.	Mean	Std. Dev.	Min	Max
Q33_1	Data about people's online behaviours (e.g. click-streams; logs; search histories...)	180	32.97	32.69	0	100
Q33_2	Data about geographical location (e.g. GPS or mobile telephone signals...)	180	33.74	25.62	0	100
Q33_3	Unstructured data like voice, text or images (e.g. blogs; tweets; footages; videos...)	180	33.30	30.28	0	100
Q33_4	Data about individuals' economic transactions (e.g. purchasing histories; credit cards operations...)	180	33.75	28.42	0	100
Q33_5	Data about people's attitudes (e.g. survey opinions; “like” buttons...)	180	31.57	24.72	0	100
Q33_6	Data about people's attributes (e.g. ethnicity; occupation; health conditions; sexual habits...)	180	26.26	17.24	0	100
	Data Protection Regulatory Regime (REG)	Obs.	Mean	Std. Dev.	Min	Max
Q47_1	Data protection law is enforced in a consistent, reliable and predictable manner.	170	3	2.96	1.73	7
Q47_2	Data protection authorities have the power and the resources to impose serious sanctions if data are processed unlawfully.	170	3	2.76	1.83	7
Q47_3	Tighter data protection regulations are necessary to ensure that all organisations meet minimum information security standards.	170	3	2.84	1.80	7

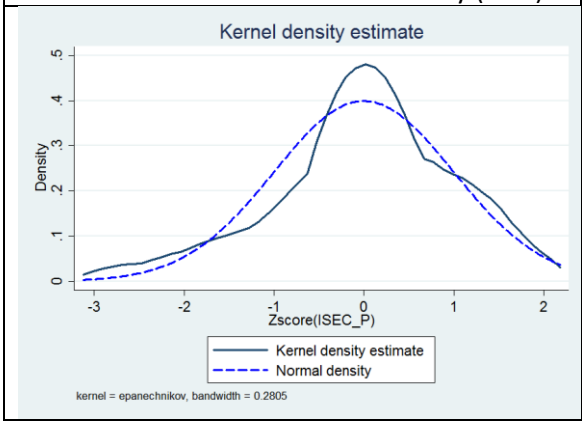
	Compliance with Data Controllers' Obligations (DPP)	Obs.	Mean	Std. Dev.	Min	Max
Q31_1	We keep data complete, accurate and up-to-date.	175	3	3.15	1.63	7
Q31_2	We try to collect the minimum amount of data necessary to fulfil a specific objective.	175	3	2.99	1.73	7
Q31_3	We delete data once the objective for which they have been collected is achieved.	175	4	3.71	1.96	7
Q31_4	We only share individuals' data with authorised third parties.	173	2	1.77	1.46	7
Q41_1	We sanction those who use or handle personal data inappropriately.	167	2	2.48	1.74	7
Q41_2	Strong security measures protect data from unauthorised use.	167	2	2.44	1.58	7
Q41_3	We have procedures in place to compensate individuals in case data were lost, manipulated or stolen.	165	4	4.05	1.88	7
	Respect of Data Subjects' Rights (DSR)	Obs.	Mean	Std. Dev.	Min	Max
Q37_1	We always obtain explicit consent from individuals before processing their data.	173	3	3.29	1.96	7
Q37_2	Individuals are fully informed about all aspects related to the processing of their data.	172	3	2.98	1.88	7
Q37_3	We can easily satisfy individuals' requests to end the processing of their data.	170	3	3.16	1.85	7
Q37_4	We have procedures in place to let the individuals rectify inaccurate data.	171	3	2.61	1.60	7
	Organisational Privacy Culture (PRV)	Obs.	Mean	Std. Dev.	Min	Max
Q35_1	Privacy represents a distinctive feature of my brand/organisation.	174	2	2.5	1.8	7
Q35_2	Remarkable human and financial resources are devoted to secure information.	174	3	3.2	1.7	7
Q35_3	Privacy is a core value, central to our organisational culture.	174	3	2.5	1.8	7
	Rationale behind Investing in Information Security (ISEC)	Obs.	Mean	Std. Dev.	Min	Max
Q43_1	To manage the risk of high litigation costs	154	6	60.7	29.7	100
Q43_2	To reflect high industry information security standards	152	5	52.9	33.6	100
Q43_3	To manage the risk of economic loss	162	7	73.5	27.5	100
Q43_4	To manage reputational risks	156	6	57.9	32.2	100
Q43_5	To improve service/product quality	147	5	51.7	31.1	100
Q43_6	To avoid costly enforcement action by regulators	157	6	63.0	30.7	100
Q43_7	To react to previous security problems	158	6	60.2	32.6	100
Q43_8	I do not know	12	1	14.8	22.4	66
	Use of Analytics across Business Functions (FUNCT)	Obs.	Mean	Std. Dev.	Min	Max
Q26_1	Analytics used to foster marketing	191	5	47.4	34.0	100
Q26_2	Analytics used to improve security	191	4	43.6	32.2	100
Q26_3	Analytics used to gain efficiency	191	5	50.2	31.1	100
Q26_4	Analytics used to better manage human resources	191	3	33.9	29.0	100
Q26_5	Analytics used to reduce financial risks	191	5	47.4	33.4	100
Q26_6	Analytics used to take better informed strategic decisions	191	5	54.9	31.6	100
Q26_7	Analytics used to offer public policy services	191	3	32.4	33.2	100

SA.2.1 Probability distributions of all composite scores

Table 51. Probability distributions of formative indicators

Analytical Sophistication (SOPH)	Dataveillance (DVEIL)
	
Compliance with Data Protection Principles (DPP)	**Privacy Culture (PRV)**
	
Respect of Data Subjects' Rights (DSR)	**Data Protection Regulatory Regime (REG)**
	

Data Pool Variety (DPOOL)	Use of Analytics across Business Functions (FUNCT)
	

Investments in Information Security (ISEC)


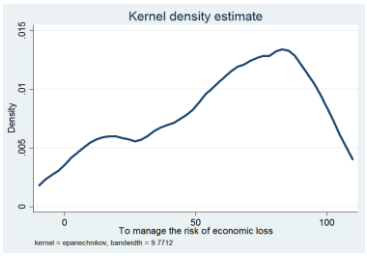
SA.2.2 Motivations behind information security investment decisions

Table 52. Probability distributions of variables measuring reasons to invest in information security

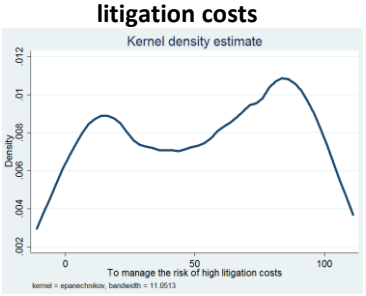
In general, what motivates investments in information security (InfoSec) inside your organisation?

Scale from 0 = "It is not at all a relevant reason to invest in InfoSec" to 100 = "It is a very relevant reason to invest in InfoSec"

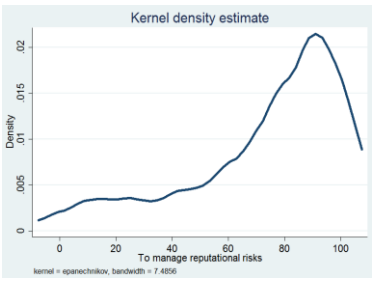
We invest in InfoSec to manage the risk of economic loss



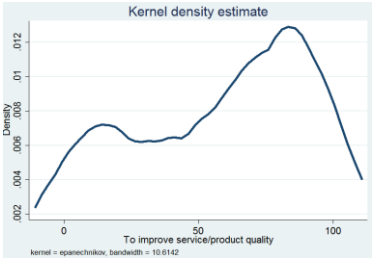
We invest in InfoSec to manage the risk of high litigation costs



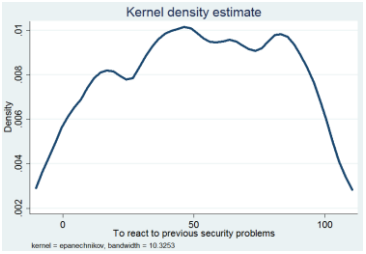
We invest in InfoSec to manage reputational risks



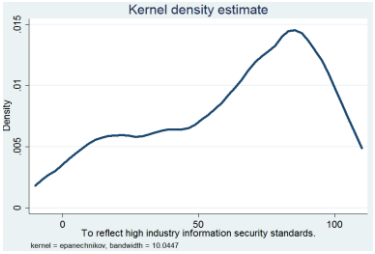
We invest in InfoSec to improve service/product quality



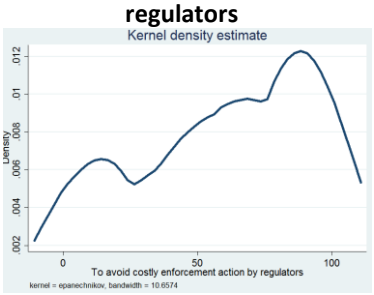
We invest in InfoSec to react to previous security problems



We invest in InfoSec to reflect high industry information security standards



We invest in InfoSec to avoid costly enforcement action by regulators



SA.2.3 Privacy and security safeguards

Table 53. Frequency of adoption of data privacy and security measures

Initiative meant to foster data privacy and security	*Freq.*	*Percent .*	*Valid Percent .*
1) Data policies that describe the rules controlling the integrity, security, quality, and use of data during its life-cycle and state change, have been adopted.	132	64.1%	82.0%
2) A Chief Privacy/Data Protection Officer is in charge of supervising all privacy-related issues.	106	51.5%	65.8%
3) The function of dealing with privacy-related matters is pursued by a designated department inside my organisation, for example the compliance office or the IT department, etc.	104	50.5%	64.6%
4) Specific policies for classifying information according to their sensitivity (e.g. secret; confidential; for internal use; etc.) are in place.	99	48.1%	61.5%
5) Employees are constantly trained to comply with privacy procedures.	99	48.1%	61.5%
6) Network and application penetration and vulnerability testing (e.g. friendly hacking).	98	47.6%	60.9%
7) Encrypted transmission of data.	97	47.1%	60.2%
8) Full-disk encryption of physical devices like laptops or PCs.	91	44.2%	56.5%
9) Workforce members are sanctioned if they do not comply with privacy procedures.	87	42.2%	54.0%
10) Consent obtained through opt-in acceptance of data processing terms and conditions.	82	39.8%	50.9%
11) Periodical external auditors' assessment of internal security standards.	70	34.0%	43.5%
12) Consent obtained through opt-out acceptance of data processing terms and conditions.	60	29.1%	37.3%
13) Certified code of practice for information security management (e.g. ISO/IEC 27002:2005).	58	28.2%	36.0%
14) Counsel of a legal firm specialized in information privacy.	55	26.7%	34.2%
15) Privacy Impact Assessments (PIAs) are undertaken.	53	25.7%	32.9%
16) Privacy Enhancing Technologies (PETs) are in use.	46	22.3%	28.6%
17) Privacy-by-design (PbD) criteria are adopted in product development.	46	22.3%	28.6%
18) Immediate notification to individuals if their data are breached, disclosed or manipulated.	44	21.4%	27.3%
19) Data breach insurance policy.	26	12.6%	16.1%
20) Binding Corporate Rules (BCRs) to manage international data transfer.	24	11.7%	14.9%
N		***206***	***161***

Table 54. Relationship between privacy and security measures part one (Phi coefficient)

CATEGORY	CPO	Privacy Dept.	Data Policy	Classified	Training	Sanctions	Friendly hacking	Disk encryption	Encrypted data transfer	Certifications
Auditor	.443	.482	.528	.481	.475	.452	.534	.559	.487	.477
BCRs	.284	.243	.298	.341	.365	.321	.288	.283	.264	.391
Certification	.397	.416	.440	.389	.516	.421	.540	.501	.393	
Classified info policy	.590	.569	.691		.604	.572	.595	.582	.570	.389
CPO		.623	.710	.590	.598	.589	.590	.597	.581	.397
Data breach insurance	.243	.203	.179	.194	.217	.171	.247	.216	.302	.163
Data Policy	.710	.699		.691	.682	.644	.660	.599	.575	.440
Disk encryption	.597	.558	.599	.582	.624	.604	.697		.688	.501
Encrypted data transfer	.581	.528	.575	.570	.579	.612	.651	.688		.393
Friendly hacking	.590	.633	.660	.595	.572	.638		.697	.651	.540
Legal firm	.343	.380	.414	.369	.365	.320	.447	.364	.316	.276
Notification	.394	.244	.367	.330	.389	.396	.330	.361	.398	.328
Opt-in	.425	.381	.525	.367	.512	.464	.400	.444	.406	.297
Opt-out	.432	.433	.453	.370	.402	.384	.389	.425	.412	.281
PbD	.399	.234	.387	.361	.417	.323	.361	.351	.385	.328
PETs	.373	.347	.367	.372	.473	.439	.435	.403	.482	.279
PIAs	.488	.410	.464	.477	.550	.469	.457	.552	.521	.374
Privacy Dept.	.623		.699	.569	.577	.598	.633	.558	.528	.416
Sanctions	.589	.598	.644	.572	.665		.638	.604	.612	.421
Training	.598	.577	.682	.604		.665	.572	.624	.579	.516

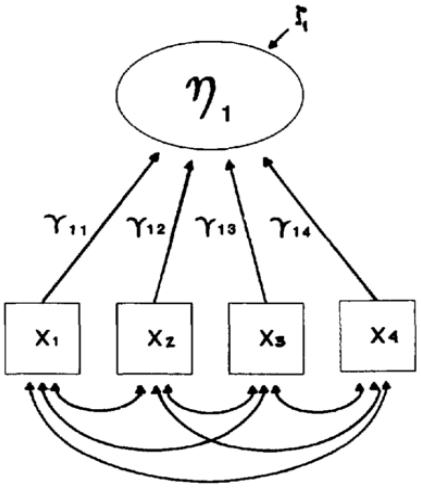
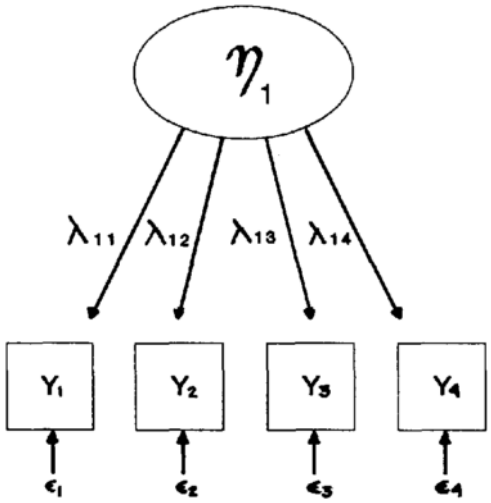
Table 55. Relationship between privacy and security measures part two (Phi coefficient)

CATEGORY	PIAs	PETs	PbD	Auditor	Notification	Opt-in	Opt-out	Legal firm	BCRs	Data breach insurance
Auditor	.331	.275	.338		.411	.325	.278	.365	.209	.198
BCRs	.226	.272	.290	.209	.307	.231	.196	.253		.067
Certification	.374	.279	.328	.477	.328	.297	.281	.276	.391	.163
Classified info policy	.477	.372	.361	.481	.330	.367	.370	.369	.341	.194
CPO	.488	.373	.399	.443	.394	.425	.432	.343	.284	.243
Data breach insurance	.152	.262	.179	.198	.228	.109	.125	.211	.067	
Data Policy	.464	.367	.387	.528	.367	.525	.453	.414	.298	.179
Disk encryption	.552	.403	.351	.559	.361	.444	.425	.364	.283	.216
Encrypted data transfer	.521	.482	.385	.487	.398	.406	.412	.316	.264	.302
Friendly hacking	.457	.435	.361	.534	.330	.400	.389	.447	.288	.247
Legal firm	.215	.253	.182	.365	.227	.290	.264		.253	.211
Notification	.335	.319	.324	.411		.403	.267	.227	.307	.228
Opt-in	.380	.316	.266	.325	.403		.321	.290	.231	.109
Opt-out	.316	.219	.268	.278	.267	.321		.264	.196	.125
PbD	.459	.563		.338	.324	.266	.268	.182	.290	.179
PETs	.461		.563	.275	.319	.316	.219	.253	.272	.262
PIAs		.461	.459	.331	.335	.380	.316	.215	.226	.152
Privacy Dept.	.410	.347	.234	.482	.244	.381	.433	.380	.243	.203
Sanctions	.469	.439	.323	.452	.396	.464	.384	.320	.321	.171
Training	.550	.473	.417	.475	.389	.512	.402	.365	.365	.217

Methodological Appendix

MA.1 Formative Vs. reflective models: A summary

Table 56. Comparison between formative and reflective measurement models

Formative Measurement Model (cause indicators)	Reflective Measurement Model (effect indicators)
a. Direction of construct–indicator causality	
The indicator *causes* the construct	The indicator is an *effect* of the construct
b. Path diagram	
 $\eta_1 = \gamma_{11}X_1 + \gamma_{12}X_2 + \gamma_{13}X_3 + \gamma_{14}X_4 + \zeta$	 $Y_1 = \lambda_{11}\eta_1 + \epsilon_1$ $Y_2 = \lambda_{12}\eta_1 + \epsilon_2$ $Y_3 = \lambda_{13}\eta_1 + \epsilon_3$ $Y_4 = \lambda_{14}\eta_1 + \epsilon_4$
c. Internal consistency perspective	
Absent. Although the cause indicators are specified to be inter-correlated, the correlations among the indicators are not relevant to the goodness of fit and conceptual viability of the model – except for the issue of multi-collinearity, which would indicate undue redundancy in the indicator set used to form the composite latent variable. In other words, the formative indicators may influence the composite construct independently of one another.	**Present.** Indicators positively associated with the same concept should be positively correlated with one another. This belief of the need for positive correlations among indicators of the same concept explains the common practice of screening correlation matrices for items that cluster together and discarding items that have near zero or negative correlations with other measures of the same construct.
d. Sampling Facets of a Construct	
Eliminating a formative indicator from the measurement model is apt to change the meaning of the composite construct; that is, because the construct is a weighted, linear combination of all its observed measures. Thus, formative indicators are not interchangeable.	There are optimal magnitudes of correlations between items. When selecting indicators of a unidimensional construct, an item should be included in a scale only if it contributes unique variance to the total scale score.

Formative Measurement Model (cause indicators)	Reflective Measurement Model (effect indicators)
e. Within-construct correlation Vs. Between-construct correlation	
Items can have any pattern of inter-correlation but should possess the same directional relationship, from the indicators to the construct.	Within-construct correlations must be greater than between-construct correlations. Correlations of indicators of the same construct should exceed the correlations between indicators from different constructs.
f. Model identification	
Identification problems are an issue in models with formative indicators. The metric of the latent composite variable can be defined either by fixing a formative indicator path to one or by fixing the factor variance to unity. Many identification problems of formative indicator constructs stem from indeterminacies associated with the construct-level error term.	Identification is made possible from information about the distribution of the observed variables. If the variables have a multi-normal distribution, then the parameters that characterize the distribution of the observed variables are the population means and the population covariance matrix. These are first- and second-order moments of the distribution. For variables that are not multi-normally distributed, higher-order moments of the distribution may help identify parameters.
g. Linear Composites as Substitutes for Latent Variables	
Error variance is represented only in the disturbance term of the latent variable, which is uncorrelated with the observed variables. Identifying the error term is not possible if the formative measurement model is estimated in isolation.	For effect indicators, the origin of the error in the composite is the measurement error in the indicators. It is possible to identify and extract observed variables' measurement error by means of factor analysis.

Source: Author's elaboration of (Bollen 1989, Bollen and Lennox 1991, Diamantopoulos and Winklhofer 2001, Coltman, Devinney et al. 2008).

MA.2 Scales treated as reflective measurement model: Validity and reliability tests

MA.2.1 Scale validity

To address the issue of measurement error, two characteristics of a measure must be considered: validity, which indicates the degree to which a measure correctly represents what it is supposed to, and reliability, which assess the extent to which the observed variable produces an accurate, error-free measure (Hair, Black et al. 2009). A reliable survey instrument is one that gets consistent results; a valid one obtains accurate results (Fink 2002). Estimates of the reliability and validity of survey questions lead to improved questionnaire design (Pfeffermann and Rao 2009).

Validity can be divided into four categories: construct, content, concurrent and criterion validity (Cronbach and Meehl 1955). Each category pursues a different objective: the goal of covering a domain corresponds to content validity, the goal of prediction to criterion validity, and the goal of measurement to construct validity (Markus and Borsboom 2013). The most difficult form of validity to assess is construct validity, as we need to establish a validation process to demonstrate that we are really measuring the construct we think we are measuring (Cronbach and Meehl 1955); as a result, the investigation of a test's construct validity "is not essentially different from the general scientific procedures for developing and confirming theories" (Cronbach and Meehl 1955: p. 300).

Validation refers to the process of investigating and documenting test validity (Markus and Borsboom 2013). Cronbach and Meehl describe a *nomological net* as a pattern of relationships between variables that partly fixed the meaning of a construct (Cronbach and Meehl 1955). Factor analysis established itself as a primary methodology for providing evidence of construct validity (Guilford 1948). Construct validity involves the acceptance of a set of operations as an adequate definition of whatever is to be measured, and must be investigated whenever no criterion or universe of content is accepted as entirely adequate to define the quality to be measured.

Content validity is established by showing that the test items are a sample of the domain in which the investigator is interested Content validity has to be established deductively, by defining a

universe of items and sampling systematically within this universe to perform the test. In criterion—also known as predictive—validity the criterion is obtained after the test is given, while concurrent validity is studied when one test is proposed as a substitute for another. Messick proposes six 'aspects' or 'components' of validity that cover a range of issues and sources of evidence to be considered in any test validation effort (Messick 1989).

Convergent evidence indicates that test scores are related to other measures of the same construct and to other variables they should relate to as predicted by the construct theory; *discriminant* evidence indicates that test scores are not unduly related to measures of other constructs (Messick 1995). Factor analysis can be used to inform evaluations of score validity. Exploratory Factor Analysis (EFA) can be used to explore the inter-items covariance structure without imposing any preconceived conditions on the outcome, while Confirmatory Factor Analysis (CFA) involves testing the fit of models to data (Thompson 2004). CFA allows the researcher to test the hypothesis that a relationship between the observed variables and their underlying latent constructs exists. The researcher uses knowledge of the theory, empirical research, or both, postulates the relationship pattern a priori and then tests the hypothesis statistically. While in EFA, all parameters implicit in a factor model must be estimated, in CFA, the researcher can 'constrain' or 'fix' certain parameters to mathematically 'permissible' values (Thompson 2004).

Initially, EFA has been used to explore the covariance structure of observable variables. In terms of the *extraction method* chosen, factor analysis has been preferred over principal components analysis, as the latter simply represents a data reduction method which does not allow to discriminate between shared and unique variance (Costello and Osborne 2005). More specifically, the *principal factor method*, an extraction technique called in SPSS 22 as 'principal axis factors' has been chosen, following recommendations made in previous studies (Fabrigar, Wegener et al. 1999), to cope with severe violations in the multivariate normality assumption. Observable variables were measured on a 1-7 bipolar or on a 0%-100% scale. EFA was performed on the

correlation, rather than on the covariance matrix to avoid problems with dealing with different measurement scales.

As reported in table 58, in all EFA models performed, the *Kaiser-Meyer-Olkin test* (KMO) test gave middling or meritorious results signalling that it was possible to proceed with the factor analysis as the null hypothesis was rejected in all EFA performed. The *Kaiser-Meyer-Olkin test* (KMO) is a measure of sampling adequacy that compares the magnitudes of the calculated correlation coefficients to the magnitudes of the partial correlation coefficients in order to indicate the proportion of variance in the observable variables which might be caused by underlying factors (Pett, Lackey et al. 2003). KMO is used to assess the suitability of the data analysed for structure detection. The test varies between 0 and 1. When evaluating the size of the overall KMO, Kaiser suggests using the following criteria for these values: a measure of above .90 can be considered "marvellous"; in the .80s is "meritorious"; in the .70s is just "middling"; and if it is less than .60-.69 is "mediocre", .50-.59 "miserable," and 0.00-0.49 is considered to be "unacceptable" (Kaiser 1974: p. 35).

The null hypothesis of the *Bartlett's Test of Sphericity* states that the correlation matrix is an identity matrix, in which all of the diagonal elements are 1 and all off diagonal elements are 0, which would indicate that variables are unrelated and therefore unsuitable for performing factor analysis (Cramer and Howitt 2004).

Table 57. KMO and Bartlett's Test of Sphericity

Model	*Constructs*	*Kaiser-Meyer-Olkin Measure of Sampling Adequacy*	*Bartlett's Test of Sphericity*			
			Approx. Chi-Square	*df*	*Sig.*	*N*
1.	DPP, ISEC	.818	858.44	91	.000	206
2.	SOPH, FUNCT	.842	1165.62	66	.000	206
3a.	DPOOL, DVEIL, PRV, DSR, REG	.771	1522.02	171	.000	206
3b.	DPOOL, DVEIL, PRV, DSR, REG	.775	1503.27	153	.000	206

In EFA, all factors with eigenvalues greater than 1.0 were initially retained. Since factor analysis is a technique that requires a large sample size and, as a rule of thumb, a bare minimum of 10

observations per variable is necessary to avoid computational difficulties (MacCallum, Widaman et al. 1999), in running EFA observable variables have been divided in three groups. Because of the limited sample size (n = 206), the researcher could not split the dataset into two datasets and run EFA on one half of cases and CFA on the other half, as suggested in other studies (Cudeck and Browne 1983). Akaike Information Criterion (AIC) was used as an alternative to conducting a split sample analysis (Bollen 1989).

The percent of total variance accounted for by each factor is reported in following table. The cumulative variance explained in the first group was 62% with three factors extracted; 70% in the second group with three factors extracted; and 70% for the third group, with five factors extracted.

Table 58. Total Variance Explained

	Group 1. DPP, ISEC			*Group 2.* SOPH, FUNCT			*Group 3b.* DPOOL, DVEIL, PRV, DSR, REG		
No of Factors	*Initial Eigenvalues*	*% of Variance*	*Cum. %*	*Initial Eigenvalues*	*% of Variance*	*Cum. %*	*Initial Eigenvalues*	*% of Variance*	*Cum. %*
1	4,83	35%	35%	4,91	41%	41%	4,26	24%	24%
2	2,50	18%	52%	2,23	19%	60%	4,00	22%	46%
3	1,31	9%	62%	1,30	11%	70%	1,75	10%	56%
4							1,55	9%	64%
5							1,03	6%	70%

No variable was dropped because of inter-item cross-loading as, when present, cross-loading values were around 0.100 and always inferior than 0.450. In the case of group 1 and group 2, three factors were extracted rather than two, as expected, for the following reasons. In group 1, questions Q21_1 and Q21_2, measuring SOPH, demanded an additional factor, despite the high correlation they show with other items of the same battery. A possible explanation is that these questions refer explicitly to the underlying information management systems, while all other questions ask specifically about analytics. In the case of DPP, two batteries of questions, not subsequently positioned in the questionnaire, were used to measure the construct. Despite the overall very high inter-item correlation, the two batteries load on two factors. Finally five factors, reflecting five constructs, were extracted in the case of group 3.

Table 59. EFA: observable variables and their underlying continua

Latent variable		*Scale*	*Observable variable*
Group 1	**SOPH:**	1-7 points	Q21_1 Q21_2 Q21_3 Q21_4 Q21_5 Q21_6
	FUNCT	0%-100%	Q26_1 Q26_2 Q26_3 Q26_4 Q26_5 Q26_6
No of cases = 206; No of variables = 12 No of factors extracted = 3			
Group 2	**DPP:**	1-7 points	Q31_1 Q31_2 Q31_3 Q31_4 Q41_1 Q41_2 Q41_3
	ISEC	0%-100%	Q43_1 Q43_2 Q43_3 Q43_4 Q43_5 Q43_6 Q43_7
No of cases = 206; No of variables = 12 No of factors extracted =3			
Group 3	**DSR:**	1-7 points	Q37_1 Q37_2 Q37_3 Q37_4
	REG:	1-7 points	Q47_1 Q47_2 Q47_3
	DVEIL:	1-7 points	Q29_1 Q29_2 Q29_3
	PRV:	1-7 points	Q35_1 Q35_2 Q35_3
	DPOOL	0%-100%	Q33_1 Q33_2 Q33_3 Q33_4 Q33_5 Q33_6
No of cases = 206; No of variables = 19 No of factors extracted = 5			

Most observable variables showed communalities higher than 0.5. Communalities, defined as the sum of squared factor loadings for the variables, indicate the proportion of each variable's variance that can be explained by the factors reflecting the underlying latent continuum. Only one variable, that is Q43_3 measuring the Data Protection Regulatory Regime (REG) construct, was excluded from the analysis because its communality was lower than 0.15. Question Q26_7 "My organisation use data analytics to offer public policy services" was not included in measuring the construct FUNC because it only referred to nonprofit and public sector organisations.

Orthogonal rotation, which refers to "a graphic visual or mathematical movement of the axes measuring the factor space" (Thompson 2004: p. 179), was used to ease the interpretation of results. As suggested in previous studies, standardised factor loadings were used before rotation (Kaiser 1958); a method called Varimax with Kaiser normalisation in SPSS 22.

Table 60. Groups 1 and 2: Rotated Factor Matrix

Group 1: Rotated Factor Matrix

		Factor		
		1	***2***	***3***
DPP	Q31_1		**.606**	.298
	Q31_2		**.853**	
	Q31_3	-.110	**.663**	
	Q31_4		**.422**	.213
	Q41_1	-.102	**.156**	**.732**

Group 2: Rotated Factor Matrix

		Factor		
		1	***2***	***3***
SOPH	Q21_1		**.196**	**.699**
	Q21_2		**.163**	**.880**
	Q21_3	-.175	**.697**	.286
	Q21_4	-.183	**.829**	
	Q21_5	-.161	**.914**	

	Q41_2	-.151	**.266**	**.848**
	Q41_3	-.194	**.119**	**.427**
ISEC	Q43_1	**.793**		
	Q43_2	**.770**		
	Q43_3	**.597**	-.162	-.316
	Q43_4	**.663**	-.165	-.247
	Q43_5	**.608**		
	Q43_6	**.704**	-.264	-.326
	Q43_7	**.747**		

	Q21_6	-.158	**.511**	.277
FUNCT	Q26_1	**.659**	-.230	
	Q26_2	**.717**		-.114
	Q26_3	**.772**	-.204	-.151
	Q26_4	**.733**	-.100	
	Q26_5	**.799**		
	Q26_6	**.743**	-.297	-.138

Table 61. Group 3b: Rotated Factor Matrix

Group 3b. Rotated Factor Matrix

		Factor				
		1	***2***	***3***	***4***	***5***
DPOOL	Q33_1	**.672**		-.199		
	Q33_2	**.562**		-.282		
	Q33_3	**.745**		-.258	-.140	
	Q33_4	**.484**			.105	-.161
	Q33_5	**.679**		-.165	.110	
	Q33_6	**.650**				
DVEIL	Q29_1	-.160		**.722**	-.155	
	Q29_2	-.230		**.864**		
	Q29_3	-.251		**.836**		
PRV	Q35_1		**.924**		.224	
	Q35_2		**.637**		.258	
	Q35_3		**.831**		.360	
DSR	Q37_1		.245	-.100	**.638**	.157
	Q37_2		.328		**.865**	
	Q37_3		.364		**.640**	
	Q37_4		.449		**.433**	
REG	Q47_1	-.103			.115	**.765**
	Q47_2					**.893**

MA.2.2 Scale reliability

Finally, each scale has been subject to internal reliability analysis, since evidence of homogeneity within the test is also relevant in judging validity. Cronbach's alpha (Cronbach 1951) has traditionally been used to estimate the internal consistency reliability of the measures (MacKenzie, Podsakoff et al. 2011). The test has been designed to measure one factor each time, and, because each scale is composed of an unweighted sum, the factor loadings are expected to all contribute roughly equal information to the score. It is possible to use a single measurement instrument administered to a group of people on one occasion to estimate reliability (Jeffrey T. Steedle 2010). Reliability coefficients are computed as ratios of reliable variance or reliable sum of squares divided by total score variance or the total sum of squares (Thompson 2004). Cronbach Alpha varies between zero and one, and a measure of at least 0.70 is considered to be acceptable (Peterson 1994).

Although a matrix of inter-correlations often points out profitable ways of dividing the construct into more meaningful parts (Guilford 1948), reliability coefficients here obtained did not provide evidence of the need of having more factors than those initially foreseen. In addition, and despite the problem with the limited representation that a two-item scale can provide of the much larger construct domain (Hulin and Cudeck 2001), in the case of the construct REG, the value of Alpha was satisfactory enough to keep the two-item solution (ALPHA = 0.829).

Table 62. Reliability test

	Cronbach's Alpha
SOPH: Q21_1 Q21_2 Q21_3 Q21_4 Q21_5 Q21_6	0.824
DVEIL: Q29_1 Q29_2 Q29_3	0.873
DPP: Q31_1 Q31_2 Q31_3 Q31_4 Q41_1 Q41_2 Q41_3	0.771
PRV: Q35_1 Q35_2 Q35_3	0.881
DSR: Q37_1 Q37_2 Q37_3 Q37_4	0.824
REG: Q47_1 Q47_2	0.829
ISEC: Q43_1 Q43_2 Q43_3 Q43_4 Q43_5 Q43_6 Q43_7	0.887
DPOOL: Q33_1 Q33_2 Q33_3 Q33_4 Q33_5 Q33_6	0.809
FUNCT: Q26_1 Q26_2 Q26_3 Q26_4 Q26_5 Q26_6	0.888

It was not possible to test whether the repeated administration of the instrument would yield similar results and then assessing test-retest reliability using canonical correlation, quantified by the intraclass correlation coefficient (Fleiss and Cohen 1973). However, split-half reliability, quantified by the Spearman-Brown prophesy formula (Brown 1910, Spearman 1910), was used to assess whether each scale was measuring different constructs. The test was performed in SPSS 22 with satisfactory results for all scales.

MA.3 NADPO members' feedback on survey items

Respondent no.	Start time	End time	Duration	Type of Org	What is your job title?	Years of working experience
1	03 Jun 2013 06:04 PM	03 Jun 2013 07:03 PM	59m 39s	Nonprofit	Complaints and Information Rights Officer	Between 15 and 20 years
2	05 Jun 201303:29 PM	05 Jun 201303:35 PM	5m 59s	Nonprofit	N/A	N/A
3	17 Jun 201305:02 PM	17 Jun 201305:28 PM	26m 5s	Nonprofit, Employment and training	Rights and Records Officer	Between 5 and 10 years
4	20 Jun 201306:54 PM	20 Jun 201307:15 PM	21m 8s	For profit	Head of data protection	Between 20 and 25 years
5	21 Jun 201310:09 AM	21 Jun 201310:21 AM	12m 24s	For profit	Group Head of Data Protection & Privacy	More than 25 years
6	25 Jun 201309:44 AM	25 Jun 201310:07 AM	23m 22s	Nonprofit, Culture and recreation	N/A	Between 1 and 5 years
7	25 Jun 201310:51 AM	25 Jun 201311:01 AM	9m 44s	Nonprofit, Health	Head of Information Governance	More than 25 years
8	25 Jun 201302:07 PM	25 Jun 201302:12 PM	5m 14s	Nonprofit, Local Authority	N/A	N/A
9	25 Jun 201310:53 PM	25 Jun 201311:07 PM	14m 46s	Nonprofit, Government agencies and public bodies	Information Governance Officer	Between 5 and 10 years
10	26 Jun 201311:07 PM	26 Jun 201311:17 PM	10m 19s	Nonprofit, Charity for specific medical issue	Information Security Manager	Between 15 and 20 years
11	27 Jun 201301:53 PM	27 Jun 201303:13 PM	1h 20m 44s	For profit	N/A	N/A

Comments

✓ 1
✓ Suggest make clear in first sentence that the GDPR is not yet an active instrument. Suggest "This Survey is on the topic of customer 'big data' analytics in the context of the *proposed* EU New General Data Protection Regulation. ✓ I found I didn't read the second and third paragraphs in blue (too much text?) ✓ I think your definition of 'Big data' needs to include the practice of profiling. ✓ If you are not careful, you may scare people from attempting the survey, if they feel they have limited knowledge of the General data protection Reg, or if they feel the GPDR is doomed not to be passed by the European Parliament / Council of Ministers in anything like the form proposed in January 2012. at first sight with disclaimer and contacts at bottom of screen, i nearly didn't read it Very comprehensive

✓ **8**
✓ (8) Local authorities will have end users, residents and other businesses and organisations as typical customers (8) Do you need so many different turnover categories? Many people working for large organisations may not know just how large their turnover is. ✓ (8) i prefer the phrase local authority
✓ **10**
✓ (10) I work for a large council. None (or, rather, lots) of the categories in Q4 describes our functions.[1] Again - "turnover"[2] is not the right word to describe a publicly-funded organisation
✓ **11**
(11) I would be surprised if anyone were to report that they did not use cookies to track online customers. [Q5. Does your company use the Internet for any of the following?] ✓ (11) Not all not-for-profits are charities who fundraise/ have 'donors' [Q5-1: My organisation organises fund-raising campaigns on Internet]
✓ **14**
✓ (14) Suggest add a "for public policy reasons (such as whether or not to provide specific public services)"[3]
✓ **16**
✓ (16) In 2nd question the example of what are monitoring devices is different - not sure this is helpful. Again, these questions are not well-suited to a large provider of statutory public services
✓ **17**
✓ (17) Q4 [sector] not suitable for local authorities which may cover a number of the options listed (17) I'm not sure what the point of Q8 is - it's nothing to do with data protection but is more a tax question. [Q7. In the next 12 months do you expect your organisation' turnover to increase, decrease, or stay roughly the same?] (17) not sure many people would know answer
✓ **19**
(19) I think you need to separate out responses for accuracy, incompatibility, and accessibility. you can't lump all 3 together. [Q9-1: Within my organisation, data are accurate, stored in compatible formats and easily accessible. Within my organisation, data are inaccurate, stored in incompatible formats and inaccessible.] (19) These questions are very subjective - so the answers will reflect the viewpoint of the respondent which may be different to that of the majority of others in that business. [Q9. Regarding the ability of your organisation to analyse and manage data effectively...] ✓ (19) last question seems wrong for local authorities [Q9-5 Our use of data analytics builds distinctive capabilities that help my organisation outperform competitors. Our use of data analytics does not build any distinctive capability that improves our competitiveness.] (19) Big heaps of unstructured data that has accumulated over the years
✓ **21**

[1] Added the category "Government agencies and public bodies".
[2] Changed "turnover" with "budget".
[3] Added suggested option to "Q10. Does your organisation use data analytics for any of the following purposes? (Please choose all that apply)"

(21) Q10 - many people may not know how the cookies are set - especially within large companies. [Q10. Does your organisation use data analytics for any of the following purposes?]
✓ **23**
(23) Again, subjective questions which may be answered differently depending on who responds within an organisation. [Q11. Concerning the extent to which data, such as customers' or end users' data, are used within your organisation to pursue your organisation's goals, could you please rate each of the following series of statements on a scale of 1 to 7, with opposing views at either end of the scale?]
✓ **25**
(25) Most won't know, and the question about analysing data about people's conversations is bordering (if not) on the illegal. [Q12. Does your organisation analyse any of the following types of data to foster decision making?]
✓ **26**
✓ (26) Possibly the key factor in investing and promoting infosec, at least in the UK public sector, is to avoid costly enforcement action by regulators (esp. the Information Commissioner).[4] This encompasses economic and reputational risk
✓ **28**
✓ 28. Is there a missing word in fifth question? ("We sanction *those* who use or handle pd inappropriately")[5] You may want this outcome, but the seventh question implies that sharing data is wrong/to be avoided. It can be permitted, lawful and beneficial (even without consent)
✓ **30**
✓ 30. I don't like these, if the intention is to identify "good" and "bad". The law is clear that personal data can and sometimes must be processed in the absence of consent (provided other condition(s) apply). Many statutory functions require the non-consensual processing of personal data. This might mean that individuals must not be informed about the processing. It can also mean that a request to end processing cannot be complied with.[6] (30) Surely, most will say that value can be derived by analysing data - without explaining how (as they won't necessarily know).
✓ **34**
✓ 34. Do you need to define what a "reportable data breach" is? Presume you mean to the Information Commissioner.[7] (34) Many respondents may not be able to answer this question objectively. [Q16. What motivates investments in information security inside your organisation?]
✓ **38**
✓ 38. These are difficult to answer for a public authority providing statutory services. It also misrepresents some provisions of the GDPR. For instance, the "explicit consent" provision has derogations. If it were an absolute requirement it would, of course, be impossible for a public authority to comply with. I think the question should make clear that many of these are qualified requirements. (38) Not sure why people would admit to sloppy data handling procedures. [Q18. Please indicate which statement better reflects the way your organisation handles personal data][8]
✓ **40**

[4] Added suggested option.
[5] Word added.
[6] Consider adding a questions asking on what method orgs rely on to comply (consent; contract..).
[7] Changed with "serious data breaches**".**
[8] He remained on the central category (the scale is functioning as expected).

✓ 40. "Counsel of a prestigious legal firm specialized in information privacy" is an odd one. It doesn't actually specify what's meant by this - does it mean "...are regularly instructed", or "...are on a retainer", or "...have been identified to be instructed in case of need"? "Prestigious" itself is a bit odd - I'd choose expertise over prestige... ✓ (40) I don't know what you mean by 'quality attribute'.[9]
✓ **42**
(42) Some organisations are required to report breaches so their systems will be better geared to logging them. [Q20. On the base of your knowledge, how often do organisations in your sector experience serious data breaches?]
✓ **43**
✓ 43. Q27. What is your educational background?[10]
✓ **48**
(48) People honestly don't know its likely impact, as the remit of the instrument has not been agreed. [Given its likely impact, has your organisation started planning for the new Regulation?]
✓ **49**
✓ 49. Large public sector organisations don't work in terms of "turnover"
✓ **53**
✓ 53. I take issue with "Although there may be some minor amendments and drafting clarifications, most privacy lawyers expect the Regulation, once adopted, to contain essentially the same key provisions as in the draft Regulation"! It might better say something like "Most privacy lawyers expect there to be major changes to data protection legislation, with many of the provisions of the draft GDPR being implemented"[11]
✓ **54**
✓ (54) why does this matter? [educational background] Why don't you ask about relevant professional qualifications - ISEB or IAPP?
✓ **56**
✓ (56) A data protection officer does not need to analyse data so I don't see the point of this question, [Do you feel you have the necessary statistical and computational skills to analyse data?]
✓ **58**
✓ (58) Presumably you will include a fair obtaining notice somewhere to explain what will happen to personal information processed by the data controller. And the fact that cookies are used on the site.
✓ **60**
(60) I think this survey needs to probe users awareness of the risks of big data - that individuals can easily be mis-identified; that prejudicial decisions can be made on false assumptions; that profiles can be wrong and unjust; that prejudicial information can be preserved indefinitely; that it is now considered impossible to render data sets anonymous, that vastly greater numbers of people will have access to significant personal data about health, finance, behaviour, education etc., all of which can inflict great damage if they leak. You might ask - how much risk do you feel there is about big data sets being used for health? education? insurance? offending? anti-social behaviour? domestic violence? mortgages? car insurance? life insurance? house insurance? recruitment and selection?

[9] Changed with “distinctive feature” [Privacy represents a *distinctive feature* of our brand/product/service.].
[10] Changed to be an open question.
[11] Change made.

(60) All the best with this initiative!

MA.4 Complete survey instrument

The Open University Business School

The Big DP Survey is on the topic of 'big data' analytics in the context of the proposed EU New General Data Protection Regulation.

- By **'big data' analytics** we mean all those information management practices, such as data gathering, integration, matching, mining and profiling, which enable an organisation to gather insights to fulfil its strategic objectives and generate revenue.
- By **New General Data Protection Regulation** we mean the comprehensive reform of the EU's 1995 Data Protection Directive proposed by the European Commission on January 25, 2012.

-> The survey will take approximately 15 minutes to complete.

-> You will have the opportunity to sign up to receive a complimentary report with the results.

Thank you for participating in the Big data Protection Study!

To know more about the *Big Data Protection Research Project* visit our website www.bigdataprotection.co.uk or contact the research project's principal investigator, Mrs Sara Degli Esposti at The Open University Business School by email sara.degliesposti@open.ac.uk or by phone +44(0)1908 655697.

Big Data Protection RESEARCH PROJECT

START NOW

BY CLICKING THE BUTTON BELOW >>

Browser Meta Info

This question will not be displayed to the recipient.

Browser	*Example: Chrome*
Version	*Example:49.0.2623.112*
Operating System	*Example:Windows NT 6.1*
Screen Resolution	*Example:1366x768*
Flash Version	*Example:21.0.0*
Java Support	
User Agent	*Example: Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/49.0.2623.112 Safari/537.36*

0% 100%

Survey Powered By Qualtrics

1. How many staff does your organisation, or major contractor, employ?

(Only one answer allowed)

- ❍ 20,000 or more
- ❍ 10,000 – 19,999
- ❍ 5,000 – 9,999
- ❍ 1,000 – 4,999
- ❍ 500 – 999
- ❍ 250 – 499
- ❍ 50 – 249
- ❍ 10 – 49
- ❍ 1 – 9

- ❍ I do not know
- ❍ I am unemployed ➔ **GO TO QUESTION 1b**

1b. How long have you been unemployed?

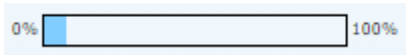
- ❍ Less than 1 month
- ❍ From 1 to 3 months
- ❍ From 3 to 6 months
- ❍ From 6 to 12 months
- ❍ From 12 to 18 months
- ❍ More than 18 months

In answering the following questions please refer to the last job you had.

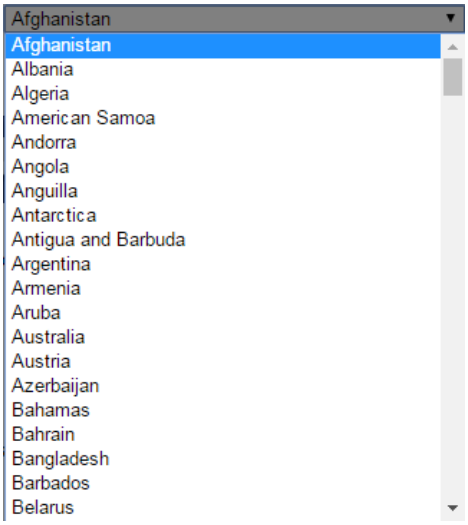
2. Is the office where you are based located in the United Kingdom?

(Only one answer allowed)

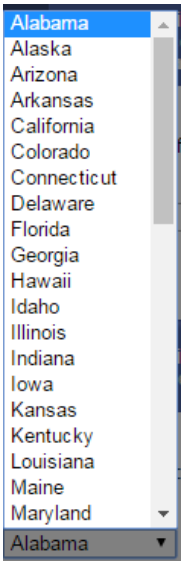
- ❍ Yes
- ❍ No ➔ **GO TO QUESTION 2b**



2b. What is the location of the office where you are based? (Please write)



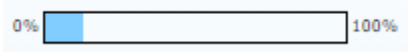
2c. Could you please specify where your office is based in the United States?



3. Is the organisation where you work a for profit or a nonprofit institution?

(Only one answer allowed)

- ❍ For profit
- ❍ Nonprofit



4. Which of the following best describes the sector in which your organisation operates?

(Only one answer allowed)

For profit organisations	**Nonprofit organisations**

- ❍ Agriculture
- ❍ Consulting / Professional Services
- ❍ Consumer Products & Retail
- ❍ Education
- ❍ Energy, Utilities & Mining
- ❍ Engineering / Construction
- ❍ Entertainment & Media
- ❍ Financial Services
- ❍ Forest / Paper / Packaging
- ❍ Government Services
- ❍ Health Industries
- ❍ Hospitality / Travel Leisure
- ❍ Industrial Manufacturing
- ❍ Technology
- ❍ Telecommunications
- ❍ Transportation & Logistics
- ❍ Other

- ❍ Community development
- ❍ Culture and recreation
- ❍ Education
- ❍ Employment and training
- ❍ Environment
- ❍ Government agencies and public bodies
- ❍ Grant-making
- ❍ Health
- ❍ Housing
- ❍ International
- ❍ Law and Advocacy
- ❍ Local authority
- ❍ Parent Teacher Associations
- ❍ Playgroups and Nurseries
- ❍ Religion
- ❍ Research
- ❍ Scout groups and youth clubs
- ❍ Social Welfare
- ❍ Umbrella bodies
- ❍ Village halls
- ❍ Other

5. What is your organisation's annual turnover/budget (please refer to 2012 revenue)?

   (Only one answer allowed)

   - ❍ £40 billion or more
   - ❍ £30 billion – £39.9 billion
   - ❍ £15 billion – £29.9 billion
   - ❍ £10 billion – £14.9 billion
   - ❍ £5 billion – £9.9 billion
   - ❍ £1 billion – £4.9 billion
   - ❍ £500 million – £999 million
   - ❍ £100 million – £499 million
   - ❍ £50 million – £99 million
   - ❍ £1 million – £49 million
   - ❍ £50,000 – £999,999
   - ❍ Less than £49.999
   - ❍ I do not know

6. Compared with 24 months ago, has your organisation's turnover increased, decreased or stayed roughly the same?

   (Only one answer allowed)

   - ❍ Turnover/Budget has increased
   - ❍ Turnover/Budget has stayed roughly the same
   - ❍ Turnover/Budget has decreased

7. In the next 12 months do you expect your organisation' turnover to increase, decrease, or stay roughly the same?

   (Only one answer allowed)

   - ❍ Turnover/Budget will increase
   - ❍ Turnover/Budget will stay roughly the same
   - ❍ Turnover/Budget will decrease

8. Taking into account all sources of income in the last financial year, did your firm generate a profit or surplus?

(Only one answer allowed)

- ❍ Yes
- ❍ No
- ❍ I do not know
- ❍ Not applicable

0% 100%

REGARDING BIG DATA..

9. CONCERNING THE ABILITY OF YOUR ORGANISATION TO ANALYSE AND MANAGE DATA EFFECTIVELY, could you please rate each of the following series of statements **on a scale of 1 to 7**, with opposing views at either end of the scale?

	1	2	3	4	5	6	7	
Within my organisation, data are accurate, stored in compatible formats and easily accessible.	❍	❍	❍	❍	❍	❍	❍	Within my organisation, data are inaccurate, stored in incompatible formats and inaccessible.
My organisation has a flexible, centralized IT infrastructure to work with data.	❍	❍	❍	❍	❍	❍	❍	My organisation lacks a flexible, centralized IT infrastructure to access and work with data.
Employees are encouraged to rely on data analytics.	❍	❍	❍	❍	❍	❍	❍	Employees are not encouraged to rely on analytics-based knowledge.
We have analysts able to mine data and get useful insights.	❍	❍	❍	❍	❍	❍	❍	We do not employ people with the necessary skills to analyse data.
Data analytics represents a distinctive, competitive capability of my organisation.	❍	❍	❍	❍	❍	❍	❍	Data analytics does not build any competitive capability within my organisation.
Digital data represents a core asset, key to our business model.	❍	❍	❍	❍	❍	❍	❍	Digital information does not add value to my organisation.

Timing

These page timer metrics will not be displayed to the recipient.

First Click	3.821 seconds
Last Click	14.046 seconds
Page Submit	0 seconds
Click Count	4 clicks

10. Do you feel you have the necessary statistical and computational skills to analyse data?

I would have no idea where to start	❍	❍	❍	❍	❍	❍	❍	I am 100% a data analyst

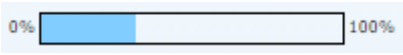
11. How knowledgeable are you about your organisation's Information Systems Management practices?

I have no idea	❍	❍	❍	❍	❍	❍	❍	I am very knowledgeable

12. Do you know **how much data** your organisation is managing in your Big data environment today?

(Only one answer allowed)

- ❍ I do not know
- ❍ 10 Terabytes or less
- ❍ 11 - 100 Terabytes
- ❍ 101 - 500 Terabytes
- ❍ 501 - 1 Pedabyte
- ❍ 2 Pedabytes or more
- ❍ None (not yet implemented a Big data environment)



13. To what extent does your organisation use **data analytics** to pursue your organisational goals in any of the following areas? Please express your opinion on a scale from 0 = "Not used" to 100 = "Definitely applied for this purpose".

To foster marketing	_____	From 0% to 100%
To improve security	_____	From 0% to 100%
To gain efficiency	_____	From 0% to 100%
To better manage human resources	_____	From 0% to 100%
To reduce financial risks	_____	From 0% to 100%
To take better informed strategic decisions	_____	From 0% to 100%
To offer public policy services	_____	From 0% to 100%

14. Does your organisation do any of the following?

(More than one answer allowed)

For profit organisations

My organisation..

- ❍ .. promotes or sells its products or services on Internet.
- ❍ .. uses monitoring devices to track customers or other people (eg web cookies, RFID, smart CCTV).
- ❍ .. generates income by storing data for other organisations.
- ❍ .. generates income by selling data.
- ❍ .. generates income by analysing data.
- ❍ .. is a ISP, hosting or cloud provider.
- ❍ .. is in the online advertising business.
- ❍ None of the above.

Nonprofit organisations

My organisation..

- ❍ .. promotes its services through a website.
- ❍ .. organises fund-raising campaigns on Internet.
- ❍ .. uses monitoring devices to track users or other people (eg web cookies, RFID, smart CCTV).
- ❍ None of the above.

1. CONCERNING THE EXTENT TO WHICH YOUR ORGANISATION COLLECTS AND PROCESSES INDIVIDUALS' DATA, SUCH AS CUSTOMERS' DATA, could you please rate each of the following series of statements on a scale of 1 to 7, with opposing views at either end of the scale?

	1	2	3	4	5	6	7	
My organisation collects data to monitor individuals' activities.	❍	❍	❍	❍	❍	❍	❍	My organisation does not monitor people's activities through data collection.
We analyse personal data to foresee and influence people's behaviour.	❍	❍	❍	❍	❍	❍	❍	Anticipating and influencing individual behaviour is not an objective of data processing.
Profiling is used to target valuable users or personalise offers.	❍	❍	❍	❍	❍	❍	❍	Individuals' data are not analysed for profiling or segmentation purposes.

Timing

These page timer metrics will not be displayed to the recipient.

First Click	3.089 seconds
Last Click	7.969 seconds
Page Submit	0 seconds
Click Count	2 clicks

2.

3. Please indicate which statement better reflects YOUR ORGANISATION'S APPROACH TO THE MANAGEMENT OF INDIVIDUALS' DATA, on a scale of 1 to 7, with opposing views at either end of the scale.

	1	2	3	4	5	6	7	
We keep data complete, accurate and up-to-date.	❍	❍	❍	❍	❍	❍	❍	We deal with partial, inaccurate and outdated data.
We try to collect the minimum amount of data necessary to fulfil a specific objective.	❍	❍	❍	❍	❍	❍	❍	We collect as much data as we can to fulfil new objectives.
We delete data once the objective for which they have been collected is achieved.	❍	❍	❍	❍	❍	❍	❍	We retain data indefinitely in case of future use.
We only share individuals' data with authorised third parties.	❍	❍	❍	❍	❍	❍	❍	We share individuals' data with any third party.

Timing

These page timer metrics will not be displayed to the recipient.

First Click 3.093 seconds
Last Click 45.201 seconds
Page Submit 0 seconds
Click Count 4 clicks

4. To what extent does your organisation analyse any of the following types of data? Please express your opinion on a scale from 0 = "Type of data not analysed" to 100 = "Type of data constantly analysed".

	Scale from 0 = "Type of data not analysed" to 100 = "Type of data constantly analysed"
Data about people's online behaviours (eg click-streams; logs; search histories..)	
Data about geographical location (eg GPS or mobile telephone signals..)	
Unstructured data like voice, text or images (eg blogs; tweets; footages; videos..)	
Data about individuals' economic transactions (eg purchasing histories; credit cards operations..)	
Data about people's attitudes (eg survey opinions; "like" buttons..)	
Data about people's attributes (eg ethnicity; occupation; health conditions; sexual habits..)	

REGARDING PRIVACY, INFORMATION SECURITY & DATA PROTECTION..

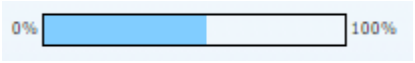
5. Concerning your organisation's approach to individuals' privacy, such as customers' privacy, and information security, could you please rate each of the following series of statements on a scale of 1 to 7, with opposing views at either end of the scale?

	1	2	3	4	5	6	7	
Privacy represents a distinctive feature of my brand/organisation.	❍	❍	❍	❍	❍	❍	❍	Privacy is not one of my brand/organisation's distinctive features.
Remarkable human and financial resources are devoted to secure information.	❍	❍	❍	❍	❍	❍	❍	Almost no human or financial resources are dedicated to information security.
Privacy is a core value, central to our organisational culture.	❍	❍	❍	❍	❍	❍	❍	Privacy does not represent an essential part of the organisational culture.

Timing

These page timer metrics will not be displayed to the recipient.

First Click	5.218 seconds
Last Click	15.692 seconds
Page Submit	0 seconds
Click Count	4 clicks



6. Please indicate which statement better reflects the way your organisation handles personal data, on a scale of 1 to 7, with opposing views at either end of the scale.

	1	2	3	4	5	6	7	
We always obtain explicit consent from individuals before processing their data.	❍	❍	❍	❍	❍	❍	❍	We do not rely on consent but on alternative means for processing personal data (eg the processing is necessary in relation to a contract).
Individuals are fully informed about all aspects related to the processing of their data.	❍	❍	❍	❍	❍	❍	❍	We do not have to inform individuals about all aspects related to the processing of their data.
We can easily satisfy individuals' requests to end the processing of their data.	❍	❍	❍	❍	❍	❍	❍	We cannot satisfy individuals' requests to stop the processing of their own data.
We have procedures in place to let the individuals rectify inaccurate data.	❍	❍	❍	❍	❍	❍	❍	There are no means for rectifying inaccurate data on the basis of an individual request.

Timing

These page timer metrics will not be displayed to the recipient.

First Click	2.294 seconds
Last Click	16.04 seconds
Page Submit	0 seconds
Click Count	6 clicks

For profit organisations	**Nonprofit organisations**

7. Who is your company's typical customer?

(Only one answer allowed)

- ❍ End-user consumers (B2C)
- ❍ Other businesses or organisations (B2B)
- ❍ The Government or other public bodies

Who is your organisation's typical end-user?

(Only one answer allowed)

- ❍ Citizens
- ❍ Other organisations
- ❍ The Government or other public bodies

8. Please indicate which statement better reflects your organisation's approach to secure personal data, on a scale of 1 to 7, with opposing views at either end of the scale.

	1	2	3	4	5	6	7	
We sanction those who use or handle personal data inappropriately.	❍	❍	❍	❍	❍	❍	❍	No internal policy foresees any sanctions for who uses personal data inappropriately.
Strong security measures protect data from unauthorised use.	❍	❍	❍	❍	❍	❍	❍	We do not have specific security measures to protect data from unauthorised use.
We have procedures in place to compensate individuals in case data were lost, manipulated or stolen.	❍	❍	❍	❍	❍	❍	❍	Individuals will get no compensation in case anything goes wrong with their data.

Timing

These page timer metrics will not be displayed to the recipient.

First Click 5.164 seconds
Last Click 11.608 seconds
Page Submit 0 seconds
Click Count 4 clicks

9. In general, what motivates investments in information security (InfoSec) inside your organisation? Please express, on a scale from 0 = "It is not at all a relevant reason to invest in InfoSec" to 100 = "It is a very relevant reason to invest in InfoSec", how relevant each factor is for your organisation

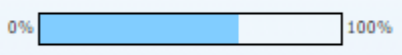
	Scale from 0 = "It is not at all a relevant reason to invest in InfoSec" to 100 = "It is a very relevant reason to invest in InfoSec"
To manage the risk of high litigation costs	
To reflect high industry information security standards	
To manage the risk of economic loss	
To manage reputational risks	
To improve service/product quality	
To avoid costly enforcement action by regulators	
To react to previous security problems	
❍ I do not know	

10. On the base of your knowledge, how often do organisations in your sector experience serious breaches of personal data?

(Only one answer allowed)

- ❍ Incidents may occur on a daily basis ➔ **GO TO QUESTION 9b**

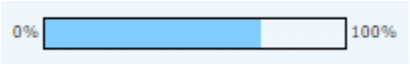
- ❍ Incidents may occur on a weekly basis ➔ **GO TO QUESTION 9b**
- ❍ Incidents may occur on a monthly basis ➔ **GO TO QUESTION 9b**
- ❍ Incidents may occur on a yearly basis
- ❍ I have never heard of any incident in my sector
- ❍ I do not know



9b. In your opinion, what most commonly causes data breaches?

(More than one answer allowed)

- ❍ Unintended disclosure (eg sensitive information posted publicly on a website or sent to the wrong party via email).
- ❍ Lost, discarded or stolen stationary electronic device (eg desktop computers, servers..).
- ❍ Hacking, malwares or spywares.
- ❍ Lost, discarded or stolen non-electronic records (eg paper documents).
- ❍ Payment Card Fraud (eg skimming devices at point-of-service terminals).
- ❍ Insiders (someone with legitimate access—such as an employee or contractor—who intentionally breaches information).
- ❍ Lost, discarded or stolen portable device (eg laptop, PDA, smart-phones, USB, CDs..).
- ❍ I do not know.
- ❍ Other (please specify):



REGARDING DATA PROTECTION LAWS..

11. Concerning data protection regulation in your country, could you please rate each of the following series of statements on a scale of 1 to 7, with opposing views at either end of the scale?

	1	2	3	4	5	6	7	
Data protection law is enforced in a consistent, reliable and predictable manner.	❍	❍	❍	❍	❍	❍	❍	Data protection law is enforced in an inconsistent, unreliable and unpredictable manner.
Data protection authorities have the power and the resources to impose serious sanctions if data are processed unlawfully.	❍	❍	❍	❍	❍	❍	❍	Data protection authorities do not have the power or resources to impose serious sanctions if data are processed unlawfully.
Tighter data protection regulations are necessary to ensure that all organisations meet minimum information security standards.	❍	❍	❍	❍	❍	❍	❍	Tighter data protection regulations are not needed to ensure organisations meet minimum information security standards.

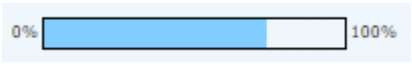
Timing

These page timer metrics will not be displayed to the recipient.

First Click	35.32 seconds
Last Click	44.129 seconds
Page Submit	0 seconds
Click Count	5 clicks

12. Do you feel you have the necessary knowledge and legal skills to understand data protection laws?

I have no idea	❍	❍	❍	❍	❍	❍	❍	I am a data protection expert



13. According to your experience and taking into account the reality of your organisation, to what extent do you consider problematic implementing each of the following provisions envisioned by the proposed new European General Data Protection Regulation?

1. Data subjects will have the right to erasure. This will allow individuals to have all personal data that business holds on them deleted or restricted. This will include all photos and any public links to, or copies of, personal data that can be found on the Internet for example in social networks or via search engines.	❍ Highly problematic ❍ Somewhat problematic ❍ Not very problematic ❍ I do not know
2. A data protection officer (DPO) must be appointed by public authorities and businesses if data of more than 5000 data subjects is processed in any consecutive 12-month period. A DPO will also have to be appointed if (i) special categories of data, (ii) location data, (iii) data relating to children, or (iv) employee data in large scale filing systems are processed.	❍ Highly problematic ❍ Somewhat problematic ❍ Not very problematic ❍ I do not know
3. Serious data breaches must be notified to both the Data Protection Agency and data subjects. Supervisory authorities will maintain a public register of	❍ Highly problematic ❍ Somewhat problematic ❍ Not very problematic

the types of breach notified. Notification must be given without undue delay.	❍ I do not know
4. Data subjects will have the right to data portability, which is a right to require a portable copy of a data subject's personal data so that they may transfer it to another data controller.	❍ Highly problematic ❍ Somewhat problematic ❍ Not very problematic ❍ I do not know
5. Consent must be given by a data subject in a clear statement or via an affirmative action (i.e. ticking a consent box when visiting a website) in cases when explicit consent would be required.	❍ Highly problematic ❍ Somewhat problematic ❍ Not very problematic ❍ I do not know
6. Data Protection Impact Assessment (PIA) must be performed annually. Companies are also encouraged to adopt Privacy by Design principles (PbD) and to certify their data processing by a supervisory authority, possibly in cooperation with accredited third party auditors.	❍ Highly problematic ❍ Somewhat problematic ❍ Not very problematic ❍ I do not know
7. The regulation will apply to organisations outside the EU whenever they process personal data of individuals in the EU. Data transfer outside the EU will be possible through Binding Corporate Rules (BCR) or in case of authorisation given by data protection authorities. Authorisations will be valid only for two years.	❍ Highly problematic ❍ Somewhat problematic ❍ Not very problematic ❍ I do not know
8. Other (please specify): _____	❍ Highly problematic ❍ Somewhat problematic ❍ Not very problematic ❍ I do not know

Timing

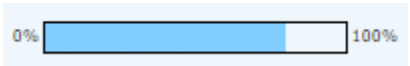
These page timer metrics will not be displayed to the recipient.

First Click	8.701 seconds
Last Click	11.638 seconds
Page Submit	0 seconds
Click Count	2 clicks

14. The draft Regulation still needs to be approved by the member states and ratified by the European Parliament before it can be adopted. It is expected that this process will take approximately two/three years. Most privacy lawyers expect there to be major changes to data protection legislation, with many of the provisions of the draft GDPR being implemented.

    Given its likely impact, has your organisation started planning for the new Regulation?

    ❍ Yes
    ❍ No
    ❍ I do not know
    ❍ Other (Please specify)



15. Which privacy or security safeguards has your organisation already adopted?

    (More than one answer allowed)

    ❍ A Chief Privacy/Data Protection Officer is in charge of supervising all privacy-related issues.
    ❍ Counsel of a legal firm specialized in information privacy.
    ❍ The function of dealing with privacy-related matters is pursued by a designated department inside my

organisation, for example the compliance office or the IT department, etc.

- ❍ Binding Corporate Rules (BCRs) to manage international data transfer.
- ❍ Data policies that describe the rules controlling the integrity, security, quality, and use of data during its life-cycle and state change, have been adopted.
- ❍ Periodical external auditors' assessment of internal security standards.
- ❍ Specific policies for classifying information according to their sensitivity (e.g. secret; confidential; for internal use; etc.) are in place.
- ❍ Immediate notification to individuals if their data are breached, disclosed or manipulated.
- ❍ Consent obtained through opt-in acceptance of data processing terms and conditions.
- ❍ Certified code of practice for information security management (e.g. ISO/IEC 27002:2005).
- ❍ Consent obtained through opt-out acceptance of data processing terms and conditions.
- ❍ Data breach insurance policy.
- ❍ Employees are constantly trained to comply with privacy procedures.
- ❍ Full-disk encryption of physical devices like laptops or PCs.
- ❍ Workforce members are sanctioned if they do not comply with privacy procedures.
- ❍ Encrypted transmission of data.
- ❍ Privacy Enhancing Technologies (PETs) are in use.
- ❍ Network and application penetration and vulnerability testing (e.g. “friendly hacking”).
- ❍ “Privacy-by-design” (PbD) criteria are adopted in product development.
- ❍ Privacy Impact Assessments (PIAs) are undertaken.
- ❍ I do not know.
- ❍ Other (please specify):

0% 100%

END OF SURVEY..

16. What is your job title?

17. Overall, how many years of working experience do you have?

(More than one answer allowed)

- ❍ Less than 1 year
- ❍ Between 1 and 5 years
- ❍ Between 5 and 10 years
- ❍ Between 10 and 15 years
- ❍ Between 15 and 20 years
- ❍ Between 20 and 25 years
- ❍ More than 25 years

18. What is your educational background?

(If you do not want to answer just leave the fields empty)

- ❍ I have a BA/BSc/MA/MSc/MBA/PhD in..

- ❍ I have Professional Certifications such as..

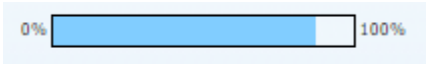
Thank you for taking part in the Big Data Protection Study!

Your input is very important to the success of the project.

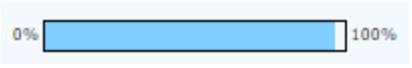
To receive the Participant's Survey Report simply type your email address in the space below. **The report will be sent soon.** Your survey responses will be kept confidential and will not be associated with the contact information provided below. That information will only be used to send you the results of the study.

We would like to discuss some issues further with you and answer your questions. Would you be interested in taking part in the follow-up of the study? If you say 'yes', our researchers will contact you for arranging an interview.

- Yes
- No



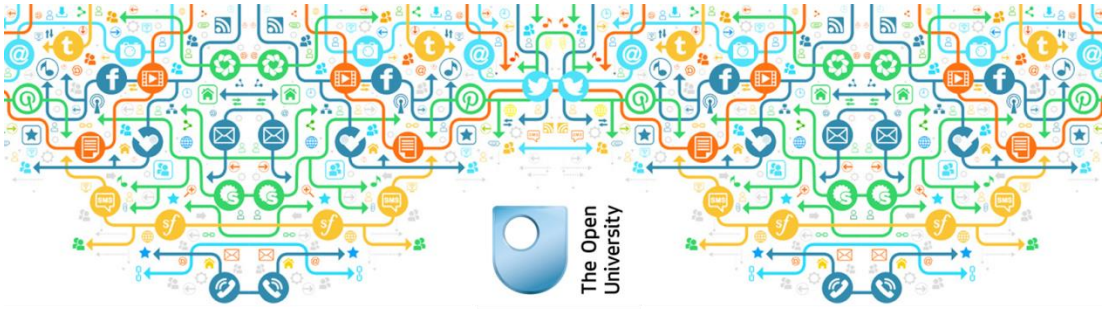
1. Any final comment?



The Open University
Business School

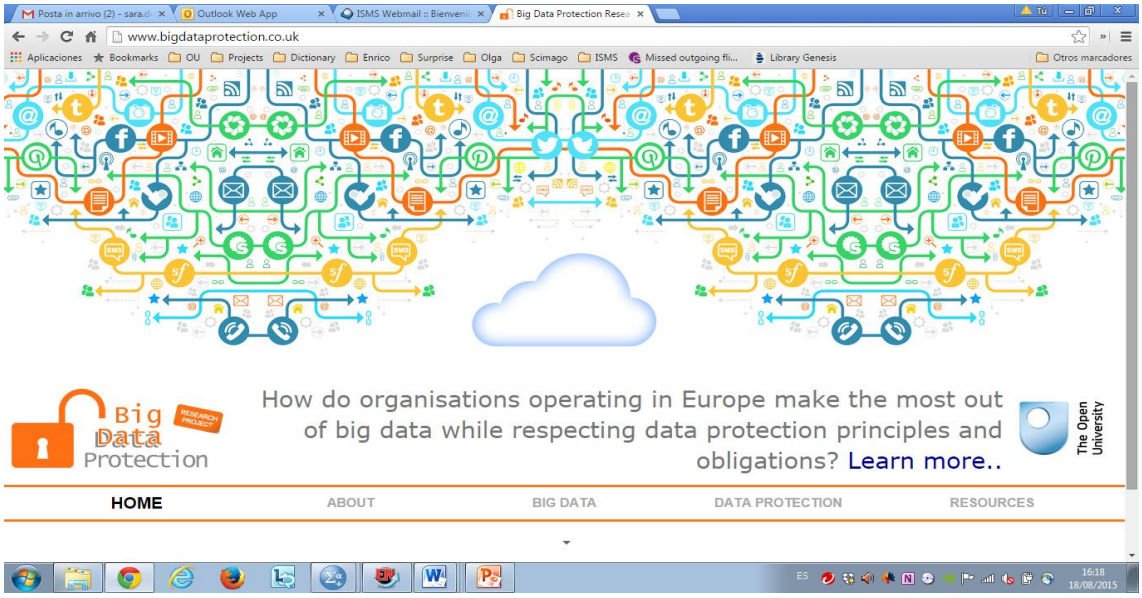
Survey Powered By Qualtrics

MA.5 Big Data Protection Study website



http://www.bigdataprotection.co.uk/

Big Data Protection
RESEARCH PROJECT



The Big Data Protection Research Project

The Big Data Protection Research Project (BigDP) studies how organisations implement data protection principles in the context of the so-called Big Data revolution. The aim is to investigate - through survey method – which information security management strategies organizations chose to extract value from personal data while respecting people's privacy. The Project looks at the data ecosystem from a data protection perspective and offers insights on how to effectively reconcile business interests with regulatory demands.

The study is conducted only for scientific purposes by researchers at The Open University Business School. No commercial or governmental organisations are involved in any way with this project. For further information please do contact the BigDP project's principal investigator, Mrs Sara Degli Esposti, either by email sara.degliesposti@open.ac.uk, phone +44(0)1908 655697 or through Twitter @survgaze.

Big Data

The term 'big data' indicates the availability of large amounts of digital information, collected through sensors, mobile devices, or other means, which can be turned into knowledge to inform people and thus offer massive opportunity for economic growth. The attribute 'big' indicates data that exceeds the processing capacity of conventional database systems. As explained by Roger Magoulas: "big data is when the size of the data becomes part of the problem". When data are too big, move too fast, or do not fit the structures of conventional database architectures, it is appropriate to talk about big data. In general, big data are measured in petabytes while 'small data' require just gigabytes (Learn more). Handling big data requires specific IT investments in order to ensure data quality, accuracy, timeliness, access to the right data, reconciliation of disparate data sources, and to guarantee compliance and information security. Big data quite often is also used as an abbreviation of 'big data analytics', which represents another way of saying data mining. Big data analytics indicates the extensive analysis - using computer algorithms and statistical and mathematical tools - of digital datasets held by large organisations, such as corporations or governments. Practices such as consumer profiling, price trends forecasting, the personalisation of offers and promotions or the availability of product recommendations on websites are all the result of putting big data analytics at work. Despite all the potential benefits coming along the big data revolution, the accumulation and analysis of large amount of data about human activities have important ethical, legal and social implications.

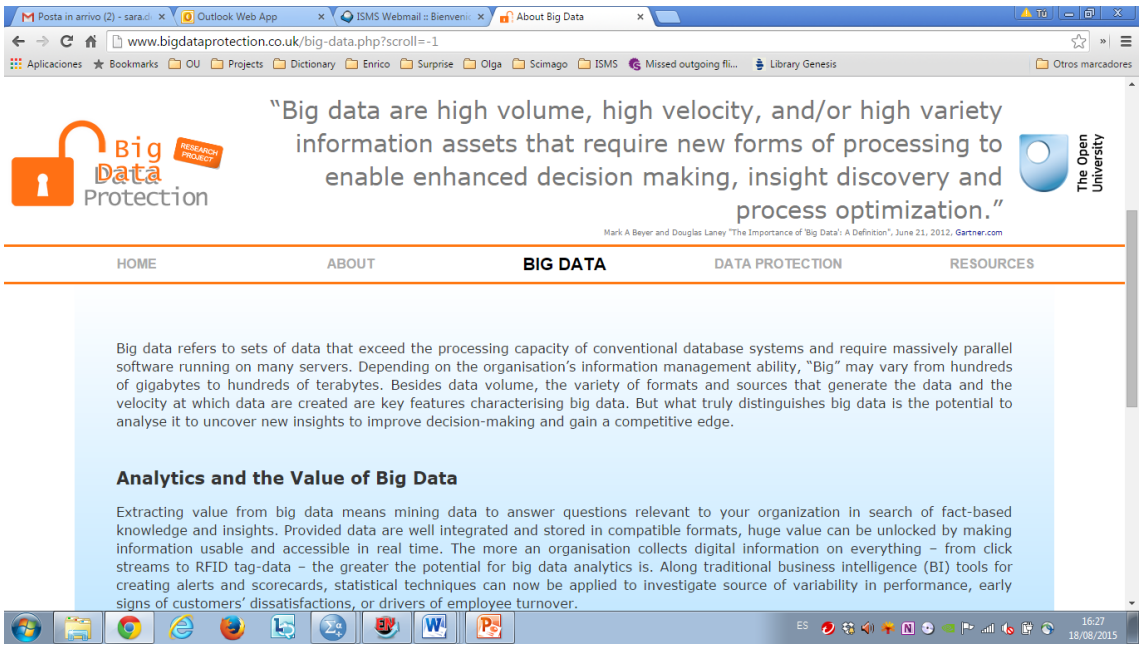
Data Protection

According to the Article 29 Data Protection Working Party despite "all its potential for innovation, big data may also pose significant risks for the protection of personal data and the right to privacy". Individual online users, who contribute to the production of 80% of the digital universe, are increasingly concerned about their information privacy as they feel to have little control over the information they create. Organisations, which manage and store 70% of all information generated by users, face the challenge of finding the right balance between data accumulation and data security and protection. Within the European Union data protection is also a fundamental human right which has been granted special safeguards. Public and private entities must comply with privacy laws such as the 1995 European Data Protection Directive and its corresponding national laws and regulations, such as the 1998 UK Data Protection Act. The proposed new EU General Data Protection Regulation might even add further obligations to data controllers/processors if approved. (Learn more)

Big Data Vs. Data Protection: emerging trade-offs

Some business commentators claim that privacy laws may hamper innovation and the creation of new business models by preventing the reuse and combination of different types of information. The 1995 EU Data Protection Directive, for example, limits the collection of personal data to the fulfilment of specific predefined purposes. It also requires the destruction of data once the purpose of which they have been collected is achieved. This provision prevents the accumulation of data, which is a necessary condition for data to become 'big'. Yet the same Directive also requires data controllers/processors to keep personal data accurate and up to date. This provision may help increase data quality, which in turns may improve estimation accuracy and big data analytics. Other commentators believe that those companies who extract value from personal data would be more willing to invest in information security and data protection, rather than the opposite. Companies who treat data as a key asset and are aware of the potential reputational and economic risks generated by data loss or unauthorised disclosure might devote more resources to data protection than organisations who have to handle personal data, but do not see any value into the data. These organisations could underestimate the importance of assigning resources to data protection programs and thus could jeopardise people's privacy. As these examples show, the relationship between big data analytics and the respect of data protection principles is not a straightforward one.

The BigDP project empirically investigates potential complementarities and conflicts emerging from the enactment of Data Protection Principles (DPPs), as stated in the 1995 Data Protection Directive, in a Big Data environment. The project will shed light on the relationship between each data protection principle and the degree of Big Data sophistication. It will offer insights on best practices and solutions and critical information security areas. The study will also contribute to the debate on the proposed New General Data Protection Regulation by offering an assessment of how problematic organisations see the implementations of the most novel and controversial provisions.



Big data refers to sets of data that exceed the processing capacity of conventional database systems and require massively parallel software running on many servers. Depending on the organisation's information management ability, "Big" may vary from hundreds of gigabytes to hundreds of terabytes. Besides data volume, the variety of formats and sources that generate the data and the velocity at which data are created are key features characterising big data. But what truly distinguishes big data is the potential to analyse it to uncover new insights to improve decision-making and gain a competitive edge.

Analytics and the Value of Big Data

Extracting value from big data means mining data to answer questions relevant to your organization in search of fact-based knowledge and insights. Provided data are well integrated and stored in compatible formats, huge value can be unlocked by making information usable and accessible in real time. The more an organisation collects digital information on everything – from click streams to RFID tag-data – the greater the potential for big data analytics is. Along traditional business intelligence (BI) tools for creating alerts and scorecards, statistical techniques can now be applied to investigate source of variability in performance, early signs of customers' dissatisfactions, or drivers of employee turnover.

Big data and analytics can create value across industries, equally in the public and in the private sector. The computer and electronic products and information sectors, as well as finance and insurance are substantially gaining from the use of big data. In the retail sector personalisation programs are fostering customer satisfaction and retention, while increasing operating margins.

Some business functions are benefiting more from the use of analytics than others. Marketing and sales departments are understanding their customers better by studying customers' behaviours (from past transactions and loyalty cards) and social data (from social media and online data) to personalise offers and promotions. But marketing is not the only area in which analytics can create a radical change. Besides identifying the most valuable customers or the most talented employee, big data and analytics can be used to optimise delivery routes, prevent out-of-stock events, minimise financial risks, stop cyber-security attacks, and so on and so forth.

The list of application is growing as the number of data crunchers and data hoarders willing to unleash the power of big data and gain a competitive edge. Yet to use analytics to outperform competitors many pieces have to follow into place. To compete on analytics an organisation must have: talented people with deep analytical skills; the capacity of integrating information from multiple data sources; C-level staff committed to the strategic use of analytics for fostering a distinctive capability. Organisations at the forefront of innovation have already got these elements and are mastering the big data game. Many more organisations have still to understand what big data and analytics mean to them, though they are already managing high volumes of data. They need help to make sense of big data and reconcile data value creation and protection within the same information management strategy.

Any public or private entity processing personal data, which means collecting, storing, analysing, disclosing or disposing data that refers to identifiable individuals, has to comply with data protection norms and principles stated in the 1995 Data Protection Directive. The Directive has been transposed into internal law by European member states. In Great Britain the 1998 Data Protection Act, which came into force on 1st March 2000, sets the norms for the lawful processing of personal data. An independent authority, the Information Commissioner's Office, is in charge of enacting and enforcing the law in the public interest, by both promoting public bodies' transparency and the respect of individuals' data privacy rights.
The following table contains a summary of those data protection principles a data controller (the person who determines the purposes for which and the manner in which any personal data are, or are to be, processed) or a data processor (any person, other than an employee of the data controller, who processes the data on behalf of the data controller) have to follow in order to process data in a fair and lawful manner.

Reform of the 1995 Data Protection Directive and draft Regulation
On the 25th of January 2012 the European Commission proposed acomprehensive reform of the EU's 1995 data protection rules with the intent of lowering administrative burdens for data controllers and processors and of strengthening individuals' privacy rights. Although the purpose and principles stated in the draft regulation show important continuity with current data protection law, the reform will bring important changes. First of all, the regulation will be immediately applicable in all European member states without the need of being enacted through national law. In addition, the regulation contains an array of new provisions. The following list gives an idea of what kind of changes the draft regulation envisions.
In terms of timing, the intention is to get a package adopted by 2014 when the European Parliament and the commission are due for re-appointment. In this case, the reform process would have taken around six years since the European Commission started its reflections on the matter. The Irish Presidency of the EU has made data protection a priority and is working hard to achieve a political agreement on the data protection reform by the end of the Irish Presidency in June 2013.

Data Protection Principles

- Fairness and Transparency

Data subjects should be fully and fairly informed about any processing of data related to them. They should be made aware of the purpose of the processing and of the identity of the entity which is processing their data.

- Legitimacy of Processing

Personal data may be lawfully processed only if (a) the data subject has unambiguously given her consent after being fully informed or at least one of the following conditions applies. Data processing is necessary because: (b) the individual wants to enter or has entered into a contract; (c) there is a legal obligation that applies to the data controller; (d) it may help safeguarding the individual's life; (e) it allows to administer justice or to exercise other public functions. Data processing may also be considered lawful if (f) the controller has some "legitimate interest" in the processing of data.

- Purpose Specification

Personal data must be collected for specified, explicit and legitimate purpose/s and not further processed in a way incompatible with those purpose/s.

- Collection Limitation

Personal data must be adequate, relevant and not excessive in relation to the purpose/s for which they are processed; and retained in a form that allows subjects' identification for no longer than needed for the purpose/s above mentioned.

- Data Quality

Personal data should be accurate, complete and kept up-to-date.

- Individual Participation

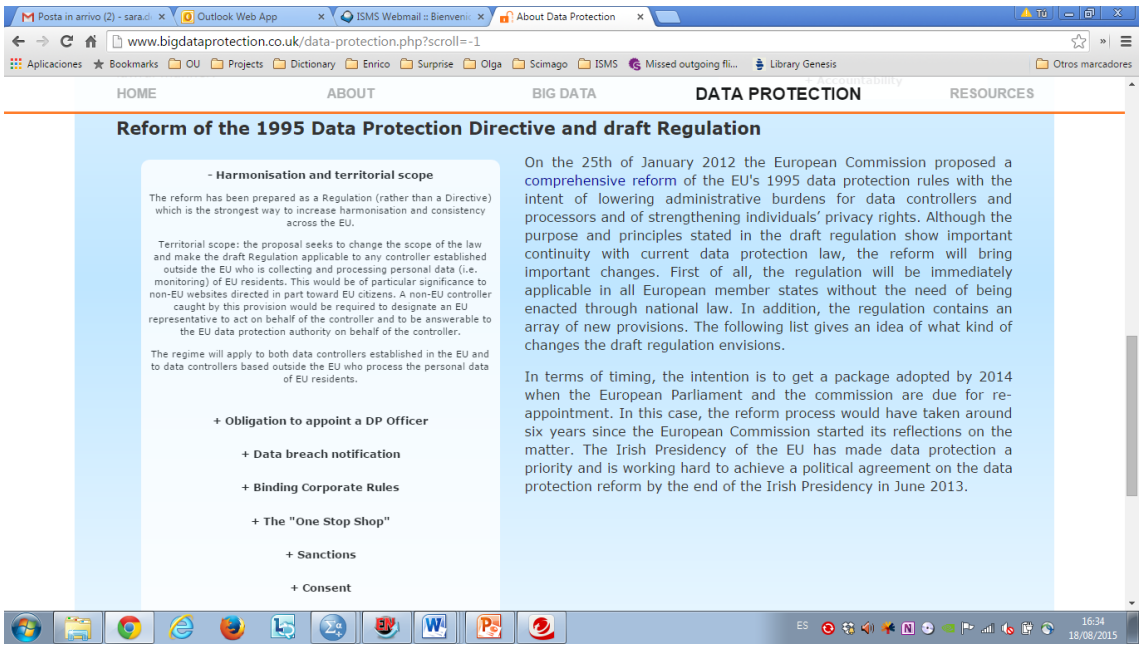
Data subjects should be allowed to easily get access to their personal records in order to rectify, erase, complete, amend, oppose or block the processing of their data.

- Security Safeguards

Appropriate technical and organizational measures should be implemented to protect personal data from unauthorised access, disclosure, modification, use, or destruction.

- Accountability

A data controller should be accountable for complying with measures which give effect to the principles stated above.



- Harmonisation and territorial scope

The reform has been prepared as a Regulation (rather than a Directive) which is the strongest way to increase harmonisation and consistency across the EU.

Territorial scope: the proposal seeks to change the scope of the law and make the draft Regulation applicable to any controller established outside the EU who is collecting and processing personal data (i.e. monitoring) of EU residents. This would be of particular significance to non-EU websites directed in part toward EU citizens. A non-EU controller caught by this provision would be required to designate an EU representative to act on behalf of the controller and to be answerable to the EU data protection authority on behalf of the controller.

The regime will apply to both data controllers established in the EU and to data controllers based outside the EU who process the personal data of EU residents.

- Obligation to appoint a DP Officer

A Data Protection Officer (DPO) will have to be appointed by all public bodies or enterprises employing more than 250 persons are processing personal data, or when the core activities of the data controller/data processor require 'regular and systematic monitoring of data subjects'. Following a review by Rapporteur Albrecht, it has now been proposed that different criteria should apply instead: where the processing of personal data "relates to more than 500 data subjects per year". This would shift the focus away from the size of the organisation and look instead at the volume of personal data being processed.

- Data breach notification

In case of data breaches, the data controller will have to notify the Data Protection Authority (DPA), where feasible within 24 hours, as well as the individuals who have been adversely affected by the data loss or disclosure. A 72-hour time frame for reporting a breach has been proposed, following a review by Rapporteur Albrecht to the European Parliament on 10 January 2013 (see DLA Piper Alert - 9 Jan 2013 for a summary of the Albrecht proposals).

- Binding Corporate Rules

The Regulation legally recognises the concept of Binding Corporate Rules for data transfer.

- The "One Stop Shop"

This concept would have a major impact on international organisations with operations across a number of EU member states. The data protection authorities in the "main establishment" of the controller would be responsible for decisions relating to the controller across its EU operations. This should offer greater harmonisation and certainty for controllers;

The data controller will act under the supervision of the national data protection authority of the EU country where the controller's main establishment is located (so called 'one-stop-shop' regulator).

Data controllers will not need to register with their local data protection authority anymore, but they will have to keep detailed records of all processing, unless there are less than 250 employees and data processing is ancillary to the controller's main activities. Data processors will have an obligation to keep documentation of all processing under their responsibility, to appoint a data protection officer if certain thresholds are met and to be liable for certain fines where there has been a breach.

- Sanctions

The regulation foresees stronger sanctions for companies processing personal data unlawfully. They range from 'up to' €250,000 to 'up to' €1,000,000. Where an enterprise is involved, they range from 'up to' 0.5% to 'up to' 2% of an 'enterprise's' annual worldwide turnover.

- Consent

All consent will have to be 'explicit', involving some affirmative action by the data subject, which means that implied consent will be no longer an option.

Parental consent will be required where a data subject is under 13 years old.

Consent given in a written declaration will have to be distinguishable from any other matters dealt with in the declaration.

Employee consent, which is currently valid in a number of EU member states, would no longer be a valid ground for processing personal data.

Direct marketing and behavioural advertising, which currently occur in a minority of member states on an implicit consent basis, are likely to require some form of explicit consent.

- Sensitive data

The definition of sensitive data is to be amended to include 'gender identity' and 'union activities'. It remains to be clarified whether this will require names and titles (eg Mr and Mrs) identifying an individual's gender to be treated as sensitive data.

- Legitimate interests

There would be a 'legitimate interests' exemption which would apply in 'exceptional circumstances'. The existing broad exemption will be removed and replaced with specific examples of where the exemption may apply. The change also clarifies that the exemption cannot be used for processing 'large amounts of personal data';

The legitimate interest condition will remain but neither third party to whom the data are disclosed nor public authorities will be allowed to rely on this clause anymore. Attention will have to be also in applying this condition if data refer to children.

Data subject will be able to object to data processing unless the controller demonstrates 'compelling legitimate grounds' for the processing which override the interests of the data subject.

Transfers outside the EEA, which are necessary for the controller's/processor's legitimate interest, provided a number of conditions have been met, will be allowed.

- Right to be forgotten

The regulation introduces a new right to be forgotten, which is the right to obtain from the data controller the erasure of personal data where, for example, the data subject withdraws consent or the data is no longer necessary and there is no legitimate reason for an organization to keep it.

- Right to data portability

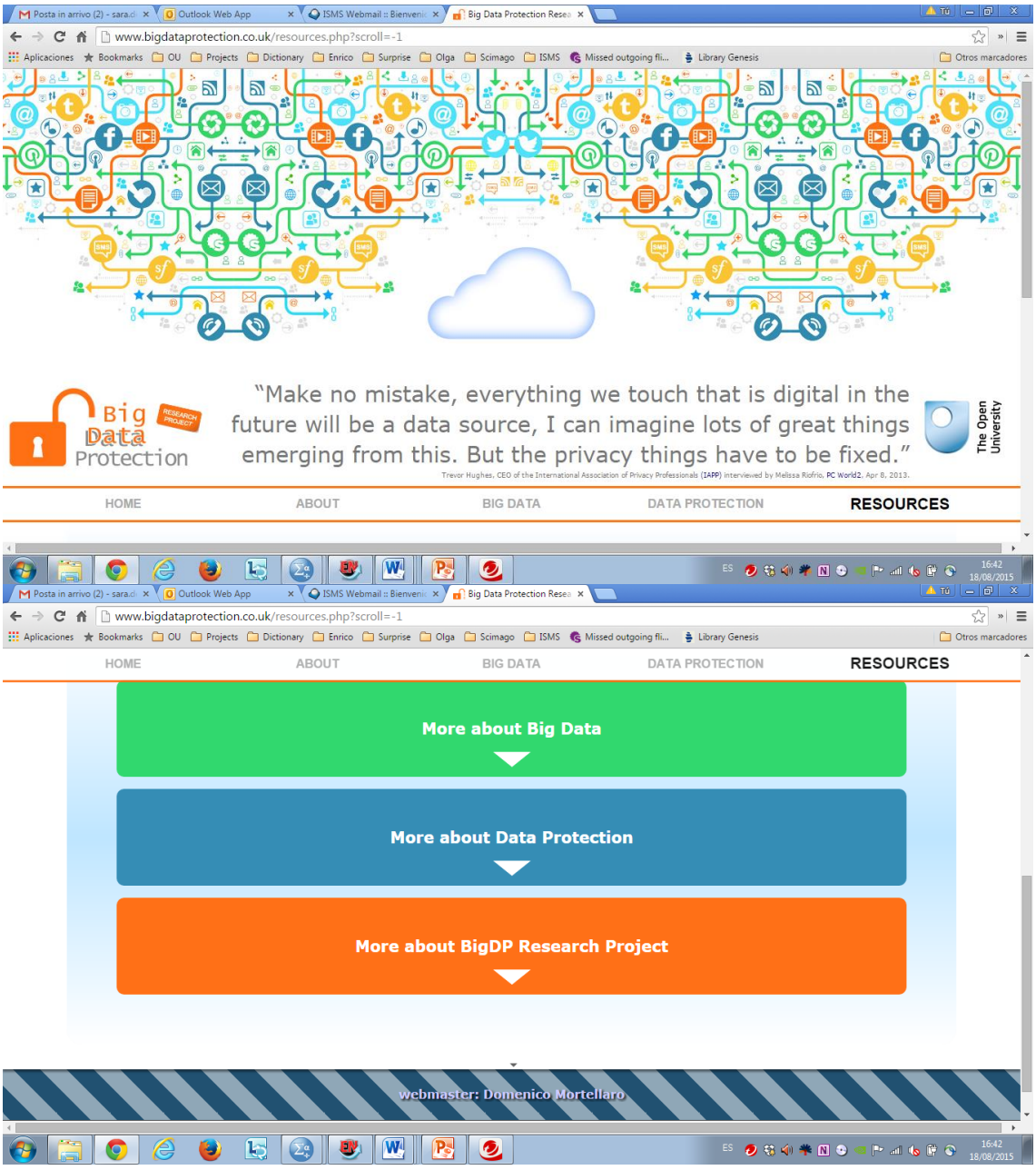
The regulation introduces a new right to data portability, which implies that a user shall be able to request a copy of personal data being processed in a format usable by this person and be able to transmit it electronically to another processing system. This provision is meant to help people transfer their personal data from one service provider to another more easily and improve competition. This right has been merged with the 'right to access' and a new principle of 'intervenability' has been introduced, putting a greater level of obligation on controllers to ensure that information is available in a commonly used electronic format;

- Privacy by design and data minimisation

This would require that (i) the controller, prior to and during the processing, implements appropriate technical and organisational measures and procedures so that the processing meets the requirements of the Regulation and ensures the protection of the data subject's rights; and (ii) the controller would need to implement mechanisms to ensure, by default, that only those personal data necessary for each specific purpose of the processing are processed, and that such data are not collected or retained beyond the minimum period necessary for those purposes.

The adoption of measures to ensure compliance and accountability, such as the use of independent auditors, adoption of 'data protection by design and default' principles, data protection impact assessments, certification mechanisms, Data Protection seals and marks, etc., are encouraged by the Regulation.

Data minimisation requirements in the current law would be expanded and would need to be incorporated into internal audit and privacy by design solutions.



MA.6 Articles published to advertise the study

MA.6.1 Blog post published in December 2013 on the ICO e-newsletter

URL: http://ico.msgfocus.com/q/1bkJRNvP4UxkK5O3bL5/wv

ico.
Information Commissioner's Office

ICO e-newsletter
December 2013

Web: www.ico.org.uk | connect with the ICO:

Your Thoughts

Join the Open University's 'Big Data Protection' Study

The Open University Business School is undertaking research into how organisations are using big data and analytics – and how they balance potential benefits with data protection considerations.

The Open University
Business School

The processing of huge digital datasets has become the new frontier for competition and innovation. More organisations than ever before are being confronted with the problem of managing large amounts of information and complying with data protection. With new European Data Protection Regulation on the horizon, how are organisations going to adapt to a regulatory change while exploiting the competitive benefits of Big Data?

If you have experience in information management or data protection, the Open University Business School would be interested in hearing your opinion. You can participate in the project by visiting the study's page on the Open University's website.

MA.6.1 Blog post published in February 2014 on the ICO e-newsletter

URL: http://ico.msgfocus.com/q/1AFxOgho8z/wv

ico.
Information Commissioner's Office

ICO e-newsletter
February 2014

Web: www.ico.org.uk | connect with the ICO:

Last chance to join the Open University's 'Big Data Protection' Study

The deadline for taking part in the Open University's 'Big Data Protection' study is fast approaching with all responses required by 28 February 2014. The study, which was originally mentioned in the December edition of our e-newsletter, is investigating how organisations are using big data and analytics and how they balance potential benefits with data protection considerations.

The Open University
Business School

You can participate in the project by visiting the study's page on the Open University website.

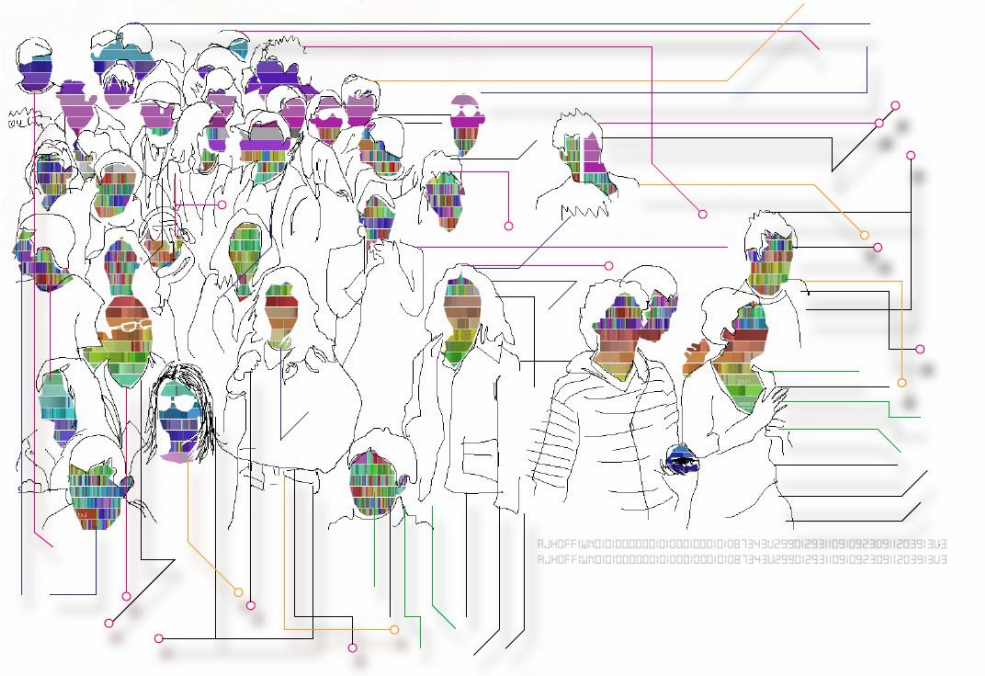
MA.6.2 Blog post published on IAPP' s Privacy Perspectives

Can We Balance Data Protection With Value Creation?

Sara Degli Esposti

Privacy Perspectives | Feb 20, 2014

URL: https://iapp.org/news/a/can-we-balance-data-protection-with-value-creation



In the last few years there has been a dramatic change in the opportunities organizations have to generate value from the data they collect about customers or service users. Customers and users are rapidly becoming collections of "data points" and organizations can learn an awful lot from the analysis of this huge accumulation of data points, also known as "Big Data."

Organizations are perhaps thrilled, dreaming about new potential applications of digital data but also a bit concerned about hidden risks and unintended consequences. Take, for example, the human rights protections placed on personal data by the EU. Regulators are watching closely, intending to preserve the eight basic privacy principles without compromising the free flow of information.

Some may ask whether it's even possible to balance the two.

Enter the Big Data Protection Project (BDPP): an Open University study on organizations' ability to leverage Big Data while complying with EU data protection principles.

The study represents a chance for you to contribute to, and learn about, the debate on the reform of the EU Data Protection Directive. It is open to staff with interests in data management or use, from all types of organizations, both for-profit and nonprofit, with interests in Europe.

The study represents a chance for you to contribute to, and learn about, the debate on the reform of the EU Data Protection Directive. It is open to staff with interests in data management or use, from all types of organizations, both for-profit and nonprofit, with interests in Europe.

Join us by visiting the study's page on the Open University website. Participants will receive a report with all the results. The BDP is a scientific project—no commercial organization is involved—with implications relevant to both policy-makers and industry representatives.

Background and Objectives of the Big Data Protection Project

What kind of legislation do we need to create that positive system of incentive for organizations to innovate in the privacy field?

There is no easy answer.

That's why we need to undertake empirical research into actual information management practices to understand the effects of regulation on people and organizations. Legal instruments conceived with the best intentions can be ineffective or detrimental in practice. However, other factors can also intervene and motivate business players to develop procedures and solutions which go far beyond compliance. Good legislation should complement market forces in bringing values and welfare to both consumers and organizations.

Is European data protection law keeping its promise of protecting users' information privacy while contributing to the flourishing of the digital economy or not? Will the proposed General Data Protection Regulation (GDPR) be able to achieve this goal? What would you suggest to do to motivate organizations to invest in information security and take information privacy seriously?

Let's consider for a second some basic ideas such as the eight fundamental data protection principles: notice, consent, purpose specification and limitation, data quality, respect of data subjects' rights, information security and accountability. Many of these ideas are present in the EU 1995 Data Protection Directive, the U.S. Fair Information Practice Principles (FIPPs) andthe 1980 OECD Guidelines. The fundamental question now is, should all these ideas be brought into the

future, as suggested in the proposed new GDPR, orshould we reconsider our approach and revise some of them, as recommended in the 21st century version of the 1980 OECD Guidelines?

A principle such as data quality, which has received very limited attention, could offer opportunities to policy-makers and businesses to reopen the debate on users' control of their personal data.

As you may know, notice and consent are often taken as examples of how very good intentions can be transformed into actions of limited importance. Rather than increase people's awareness of the growing data economy, notice and consent have produced a tick-box tendency accompanied by long and unintelligible privacy policies. Besides, consent is rarely freely granted. Individuals give their consent in exchange for some product or service or as part of a job relationship. The imbalance between the two goods traded—think about how youngsters perceive not having access to some social media as a form of social exclusion—and the lack of feasible alternatives often make an instrument, such as the current use made of consent, meaningless.

On the other hand, a principle such as data quality, which has received very limited attention, could offer opportunities to policy-makers and businesses to reopen the debate on users' control of their personal data. Having updated, accurate data is something very valuable for organizations. Data quality is also key to the success of many business models. New partnerships between users and organizations could be envisioned under this principle.

Finally, data collection limitation and purpose specification could be other examples of the divide between theory and practice: The tendency we see is that people and businesses want to share, merge and reuse data over time and to do new and unexpected things. Of course, we all want to avoid function creep and prevent any detrimental use of our personal data. We probably need new, stronger mechanisms to ensure data are used for good purposes.

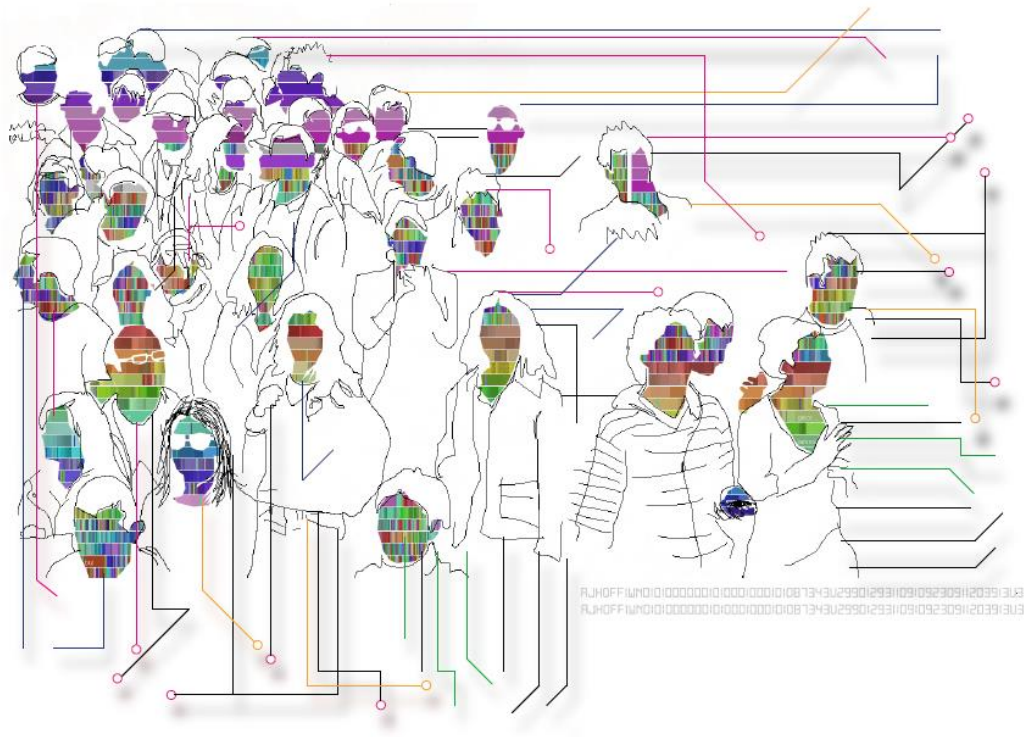
Digital data have become economic assets these days. We need good legislation to stop the black market for personal data and open the debate on how each of us wants to contribute to, and benefit from, the data economy.

Big Data Protection

RESEARCH PROJECT

Study Report

Sara Degli-Esposti, MSc



The Open University

April 2015

About the BigDP Project

The Big Data Protection project (www.bigdataprotection.co.uk) has been sponsored by the Open University in collaboration with 'The New Transparency Project: Surveillance and Social Sorting', funded by the *Social Sciences and Humanities Research Council* of Canada. The project has been developed by Sara Degli-Esposti as part of her PhD in Management at the Open University Business School. For more information please contact the author by email sara.degliesposti@open.ac.uk or Twitter @survgaze.

Acknowledgements

The author would like to thank those who contributed to give visibility to the study as well as to invite professionals to participate in the study: Carl Wiper from the *Information Commissioner's Office* (ICO); Chris Tiernan and Jon Hall of *BCS Effective Leadership in IT* (ELITE) *Group*; Omer Tene and Jedidiah Bracy of the *International Association of Privacy Professionals* (IAPP); Massimo Attoresi of the *European Data Protection Supervisor* (EDPS) authority; Ricard Martínez Martínez of *Asociación Profesional Española de Privacidad* (APEP); Llorenç Pagés Casas of *Asociación de técnicos de informática* (ATI); Stewart Dresner and Laura Linkomies of *Privacy Laws & Business*; Gavin Blackett of the Operation Research (OR) Society. She would also like to express her gratitude to her faculty Professor Kirstie Ball, Professor Elizabeth Daniel and Maureen Meadows for their valuable guidance and support with completion of this project.

Credits

The image on the cover, whose title is "Data People", was created by *andreslopz* for the BigDP project, for more information please visit the artist's website at www.andreslopz.com or send an email to info@andreslopz.com.

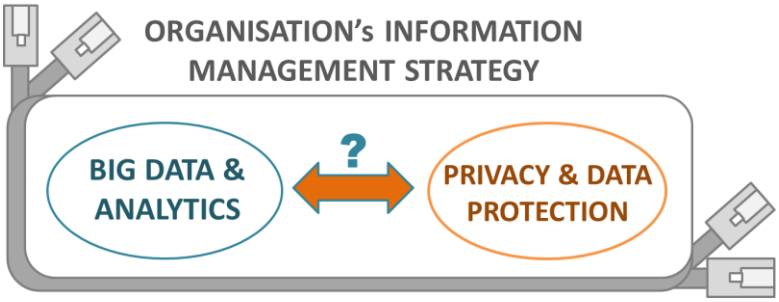
Executive Summary

The Big Data Protection project explores the way firms are using big data and analytics - and how they balance improved business performance with data protection considerations. In light of the debate around the proposed General Data Protection Regulation, the study explores the information management implications of complying with the principles stated in the 1995 Data Protection Directive at a critical time for organisations: a time when, to compete in the marketplace, businesses increasingly need to collect and analyse as much data as they can.

The study tries to answer questions such as: is it still possible, in the era of big data, to comply with data protection principles? What factors play a major role in transforming the respect of people's privacy into a key organisational value? What tools and practices should organisations adopt to strike the balance between data protection and full data usage? And, can regulation help foster innovation and high data protection and security standards? Based on survey data gathered between December 2013 and May 2014, the project hopes to offer insights on the state of compliance with data protection law and on the main drivers pushing organisations to invest in information security and privacy-protective measures.

According to the results of this study, countries where national data protection authorities have the power, and the resources, to enforce data protection laws in a consistent and predictable manner represent a positive institutional environment for organisations. In these contexts, organisations are more likely to develop a strong privacy culture, which is a necessary condition to adopt fair information practices and respect data subject's rights. Moreover, a reliable regulatory regime and a strong privacy culture help organisations use analytics to achieve target objectives and generate value. In contrast, data accumulation as an end in itself jeopardises an organisation's effort to build its privacy culture and to embed data protection into all organisational information management procedures.

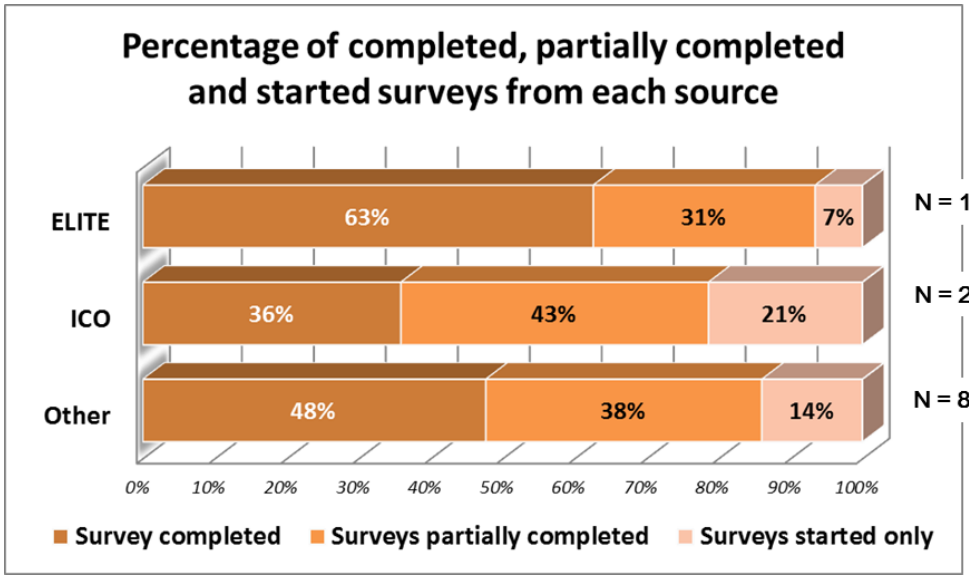
Analytics and data protection are thus compatible. Investing in an organisation's information management system, in big data analytics and in fostering a privacy-respectful organisational culture help organisations compete in the marketplace, comply with data protection obligations and respect individual privacy rights.



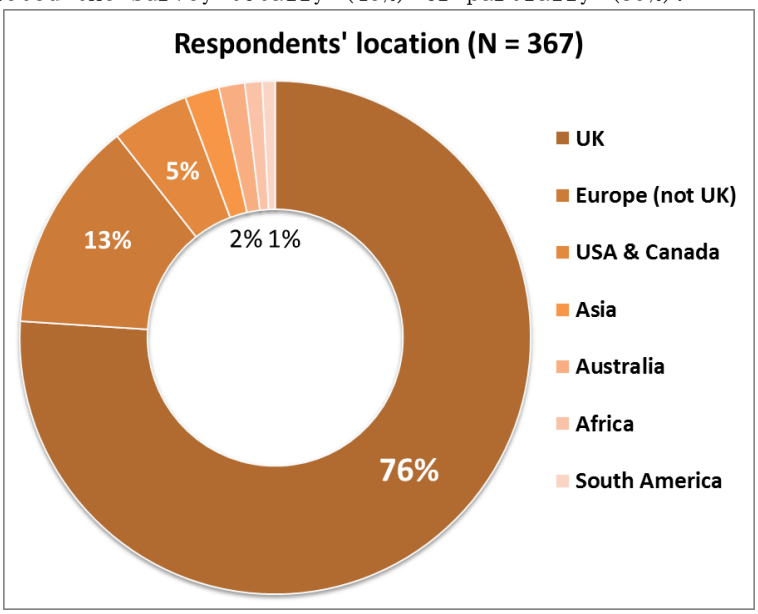
About the Data

Between January and April 2014, 441 professionals participated in the Big Data Protection Study by filling in an online survey. The UK Information Commissioner's Office (ICO) published two posts in its e-newsletter [1] advertising the study. Several specialised media, such as the magazine "Privacy Laws and Business", "Inside O.R." [2], and the online blog IAPP's "Privacy Perspectives", published articles about the study. The BCS Effective Leadership in IT (ELITE) Group [3] invited its members to participate. The European Data Protection Supervisor, the Spanish Association of Privacy Professionals (APEP), and the Spanish Association of IT professionals (ATI) also contributed

by distributing invitations to participate in the study.



84% of participants completed the survey totally (45%) or partially (39%). The number of usable surveys was thus 369. In the charts the number in parentheses (n =) indicates the total number of respondents who answered each question. In terms of participants' characteristics, respondents were mostly based in the UK (76%; n = 367), or other European countries (13%).

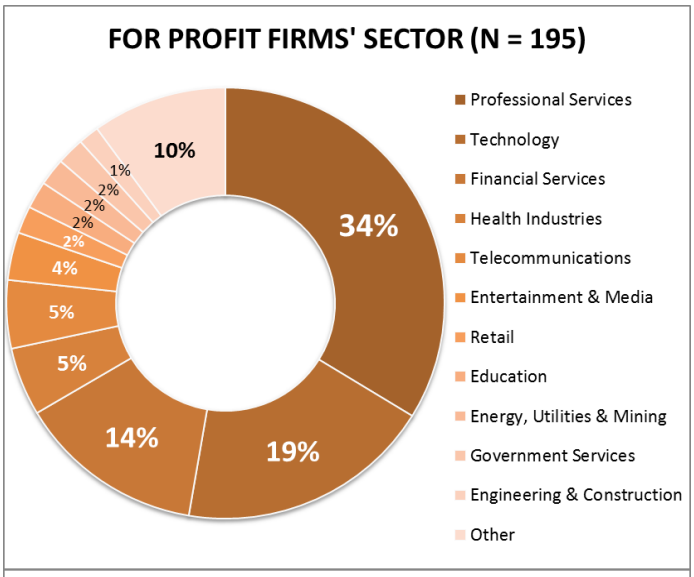


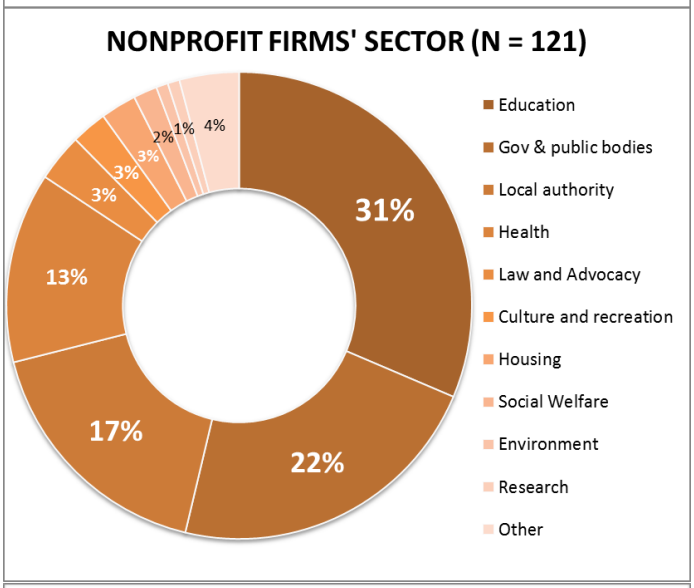
Organisational Characteristics

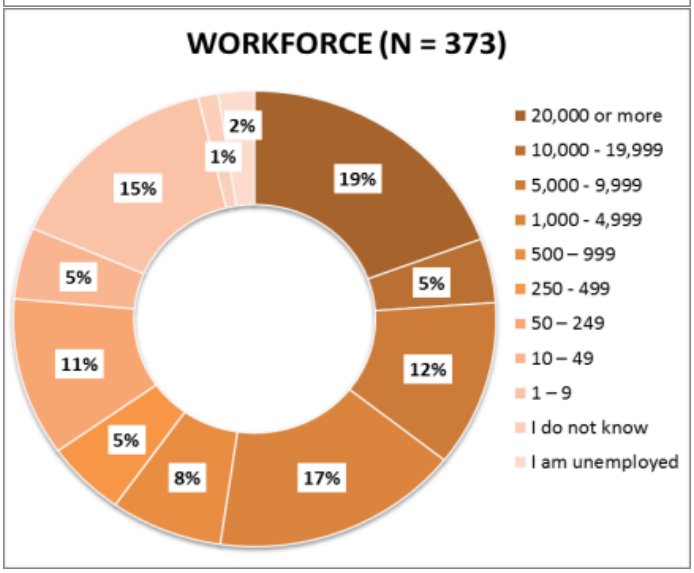
Professionals participating in the study mostly worked in the private sector (61%; n = 369), though nonprofit and public organisation were represented too (39%). Organisations of all sizes were represented in the sample.

Of the private firms' respondents, some offered consulting and professional services (18%; n = 369), or financial services (7%), while others worked for technology companies (10%).

Several other industries, such as the health sector (3%), telecommunication (3%), entertainment and media (2%), and many others, were also represented. 7% of firms surveyed were ISP, hosting or cloud providers, while another 5% were in the online advertising business.



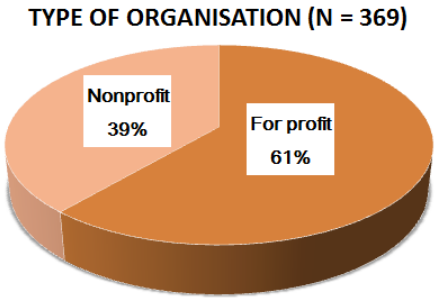


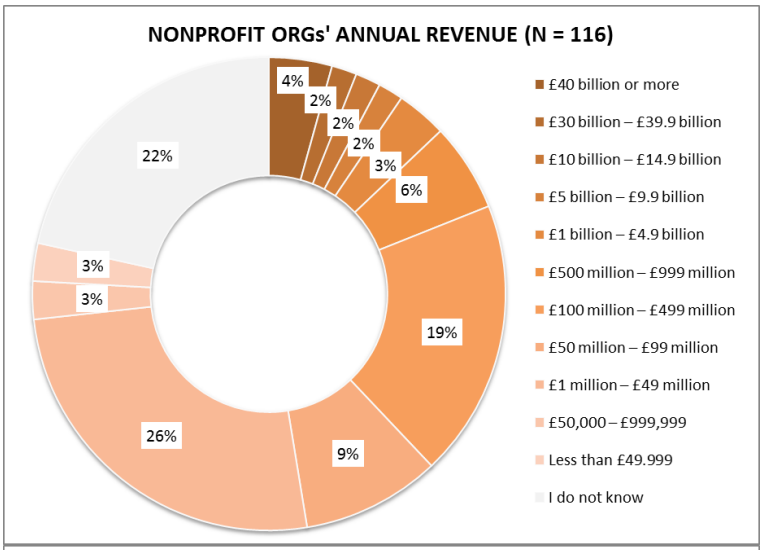


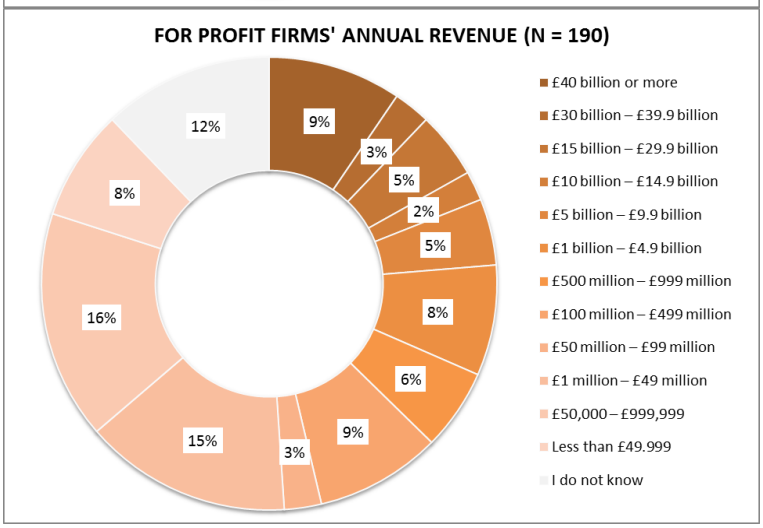
With regard to nonprofit organisations (39%; n= 369), most of them were government agencies and local authorities (13%), or organisations operating in the health sector (4%).

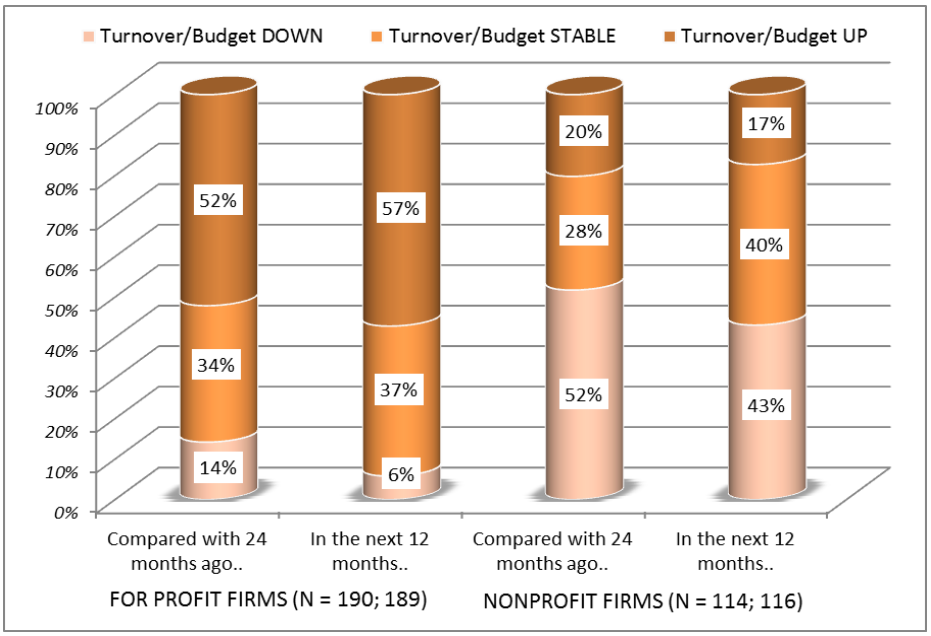
In terms of firms' annual turnover, the sample includes responses from companies with a turnover of less than £50,000 up to those with a turnover greater than £40 billion.

In terms of economic performance, while in the previous year respondents said that the income of nonprofit organisations went down (52%; n = 114), revenues of for profit organisations participating in the study not only went up (52%; n = 190), but in most cases were expected to keep growing in 2015 (57%; n = 189).

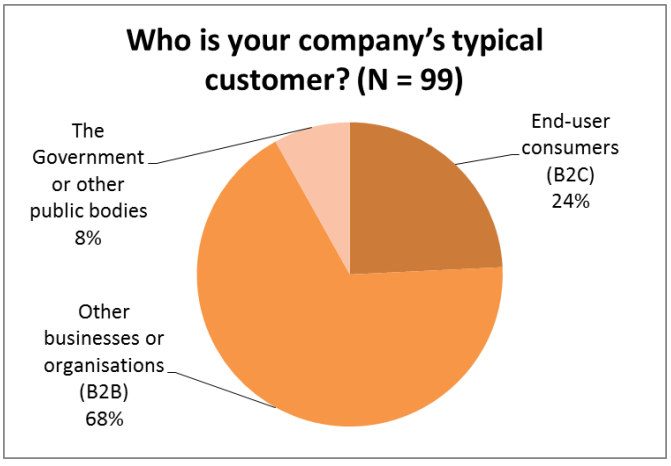


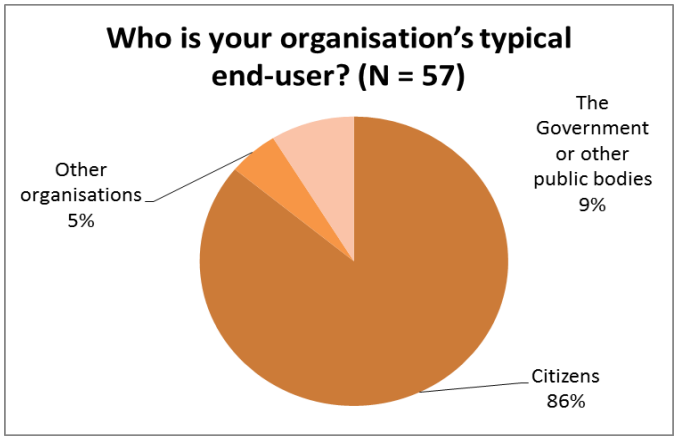






Private companies in the sample mostly worked with other organisations (B2B 68%; n = 99). In contrast, citizens were the most typical customer of nonprofit organisations present in the sample (86%; n = 57).

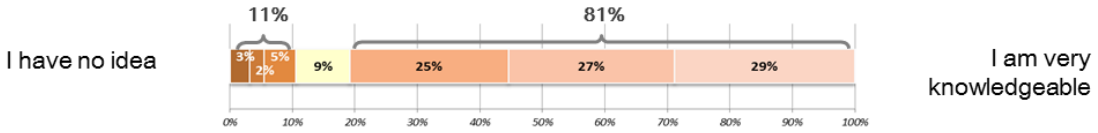




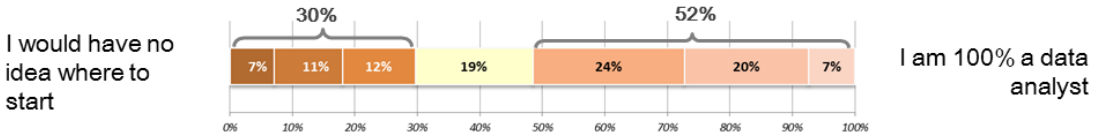
Respondents' Characteristics

Study participants were fairly knowledgeable about their organisations' information management practices (81%; n = 256). Most of them had also good analytical skills (52%; n = 256), and almost all of them said they had the necessary knowledge and legal skills to understand data protection laws (85%; n = 170).

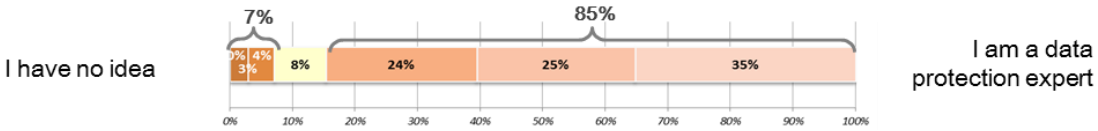
How knowledgeable are you about your organisation's Information Systems Management practices? (N = 256)



Do you feel you have the necessary statistical and computational skills to analyse data? (N = 256)

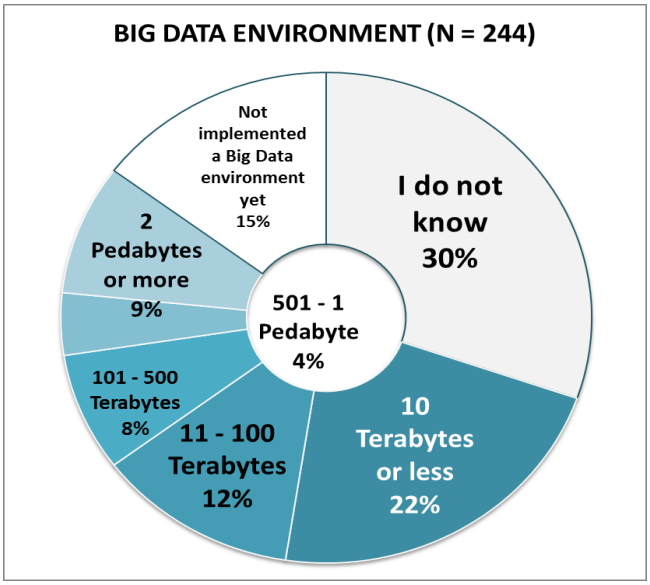


Do you feel you have the necessary knowledge and legal skills to understand data protection laws? (N = 170)

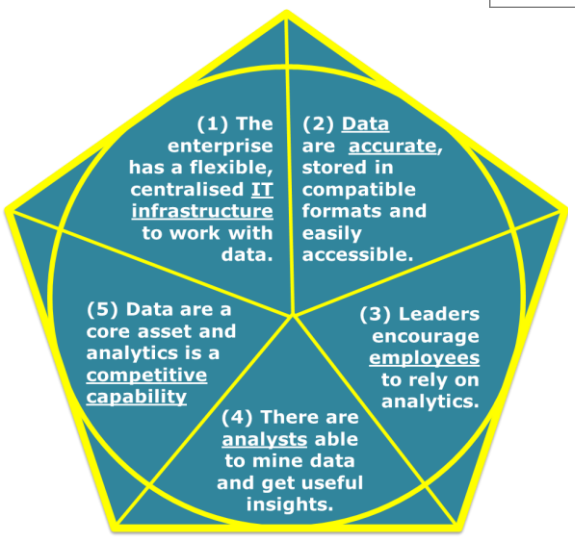


Big Data and Analytics Landscape

While Google processes 20 petabytes of data each day, most of the companies interviewed in this study deal, in general, with less than 1 Petabyte (46%; n = 244). Although most respondents said they were fairly knowledgeable about their organisation's Information Systems Management practices, 30% of participants said they had no idea about the amount of data processed in their organisations' big data environments.



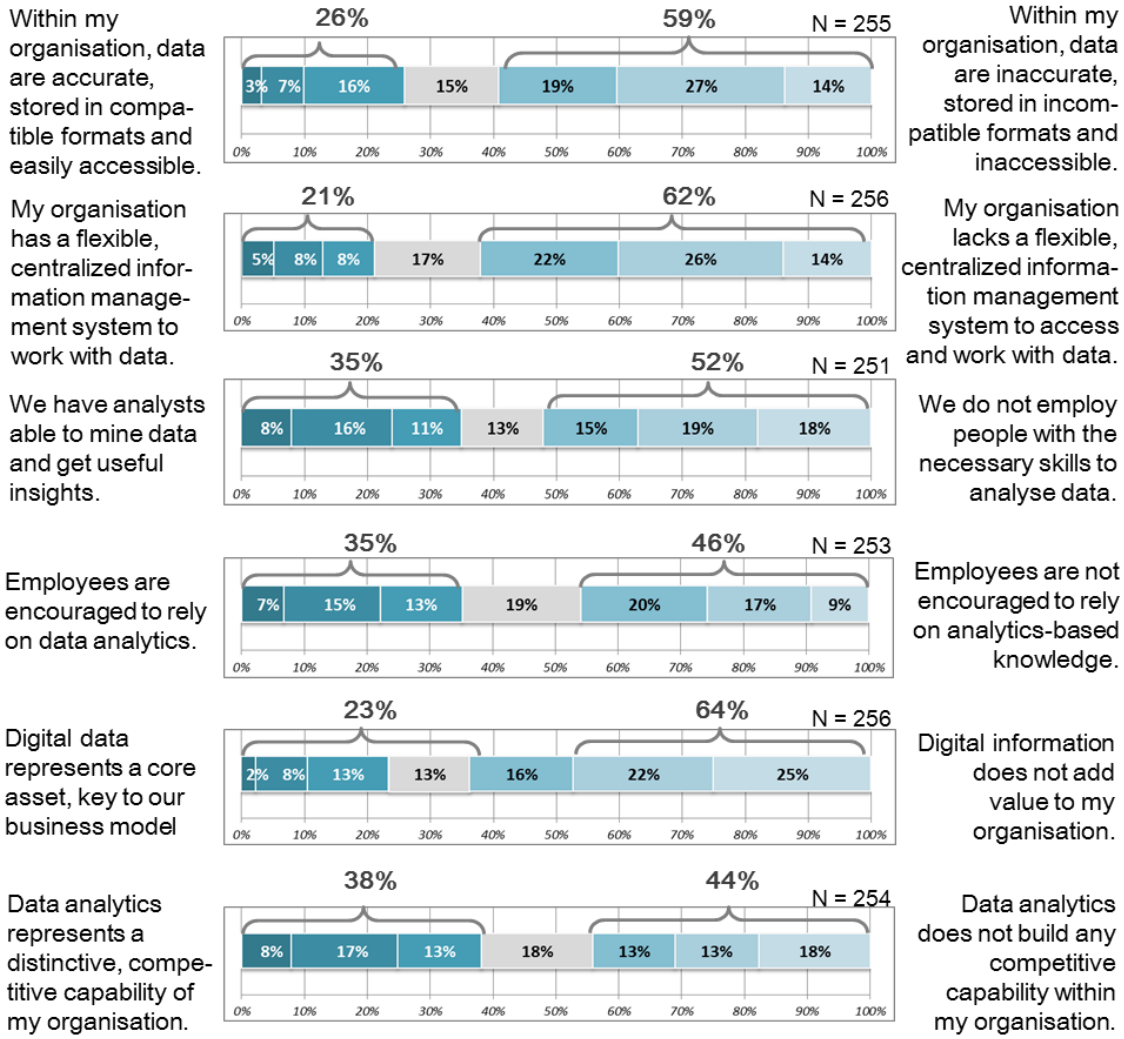
'Big data' is usually said to be all about the IT challenges of processing large volumes of fast-moving data, in different formats. Although effective data management represents a necessary condition to take advantage of big data, analytics is the key engine which triggers the creation of economic value. Big data analytics, in other words, the ability to extract actionable insights from raw data, is in fact the driver of the big data hyperbole.



5 Analytical Sophistication Constitutive Elements

But, to fully exploit the potential of big data, organisations need to become analytically sophisticated. To achieve this purpose, organisations have to embrace a specific mix of technologies, procedures and organisational values. Analytically sophisticated organisations feature in fact five basic elements [4]. They have (1) a flexible and integrated IT infrastructure that allows people to access and work with (2) high-quality data. (3) Employees are also encouraged to use analytics, and the organisation (4) employs people with the necessary statistical and computational skills to analyse data. Finally, (5) analytics helps build a key organisational competitive capability and data are considered a core asset within the organisation. In order to measure an organisation's degree of analytical sophistication the following scale has been used.

Concerning the ability of your organisation to analyse and manage data effectively, could you please rate each of the following series of statements on a scale of 1 to 7, with opposing views at either end of the scale?



By looking at the charts, it seems that most organisations still face serious problems in their attempt to become analytically sophisticated. 59% of respondents admitted they deal with inaccurate data, often stored in incompatible formats (n = 255); while 62% said their organisations do not have an adequate information management system to work with data. In addition, most respondents said that their organisations do not employ analysts (52%), and that employees were not encouraged to use analytics (46%). As a result, digital information seems not to be exploited by most organisations (64%). Nonetheless, there are a few organisations which see analytics as their distinctive, competitive capability (38%; n = 254).

Achieving Targeted Objectives through Analytics

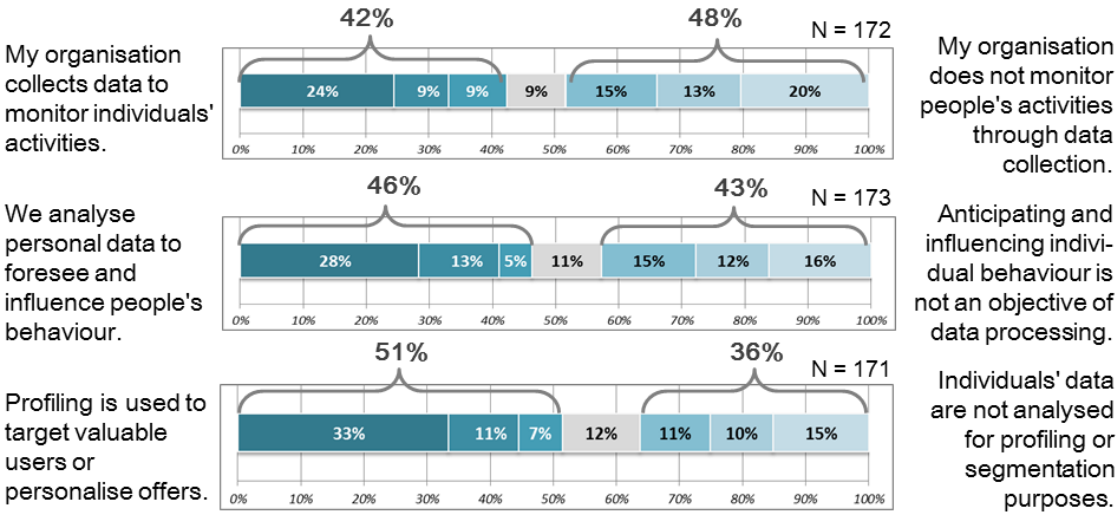
Analytics can be used to achieve several different objectives and can be applied to any organisational functional area: from customer retention or market penetration, to efficiency gains or security threat prevention. Not all organisations apply analytics across all business functions, though. Since analysts are a scarce and costly resource, analytics is usually applied to improve the performance of the most critical area within an organisation, or to improve strategic decision-making at C-level. Depending on the firm's business model, the selected functional area, which receives more attention, may vary. According to the results of this study, most organisations rely on analytics to take strategic decisions (64%; n = 189), gain efficiency (53%), reduce financial risks (51%), or improve security (46%). Private companies primarily use analytics to foster marketing (56%; n = 121). In contrast, nonprofit organisations, such as government agencies, besides using analytics to offer public policy services (44%; n = 68), seem to use analytics mostly to manage human resources (37%).

		For profit			Nonprofit		
	Analytics used to achieve the following objectives	*Not applied for this purpose*	*Partially applied for this purpose*	*Definitely applied for this purpose*	*Not applied for this purpose*	*Partially applied for this purpose*	*Definitely applied for this purpose*
1.	*To take better informed strategic decisions*	9%	25%	66%	9%	31%	60%
2.	*To foster marketing*	10%	34%	56%	16%	47%	37%
3.	*To gain efficiency*	10%	36%	54%	7%	41%	51%
4.	*To reduce financial risks*	12%	35%	54%	10%	44%	46%
5.	*To improve security*	10%	40%	50%	9%	53%	38%
6.	*To better manage human resources*	14%	52%	34%	7%	56%	37%
7.	*To offer public policy services*	NA	NA	NA	9%	47%	44%
	For Profit organisation n = 121; nonprofit organisations n = 68; total n = 189.						

Dataveillance and Targeted Analytics

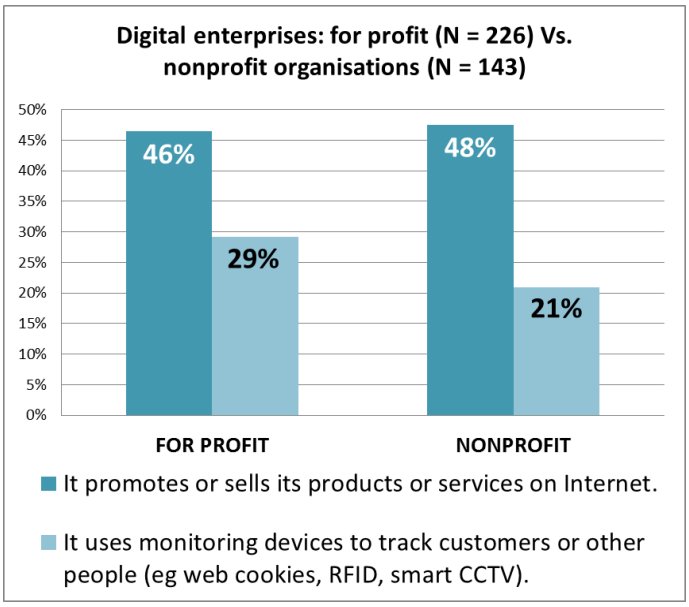
Digital data which are linked to human activities are especially valuable as they help organisations understand people's attitudes, intentions, and behaviour. Thanks to this knowledge organisations can fine tune promotions, personalise offers or improve their employee retention schemes. Within this study we call "dataveillance" the set of practices organisations use to foresee and nudge individual behaviour. Dataveillance, defined as the systematic monitoring of people or groups, by means of personal data systems, in order to regulate or govern their behaviour [5], has been measured by asking the following questions.

Concerning the extent to which your organisation collects and processes individuals' data, such as customers' data, could you please rate each of the following series of statements on a scale of 1 to 7, with opposing views at either end of the scale?



Type of Data Analysed

Huge amounts of both online and offline data are becoming increasingly available: most organisations go online to promote their initiatives (46% for profit; 48% nonprofit) and several of them rely on tracking devices to monitor potential customers or users (29% for profit; 21% nonprofit).



In terms of the type of data most commonly analysed within organisations, the public sector seems to rely more on data about

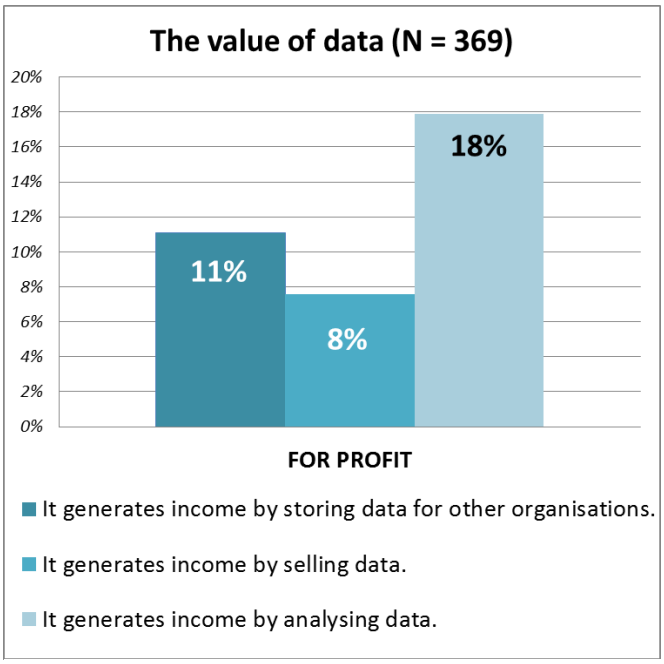
people's attributes (49%), - which may come from data collected when people request services or from census data - and on public opinion pool data (37%). The private sector tends to use data produced as a by-product of other activities, such as people's online footprints (30%) and transactional data (30%), but also opinion pool data (30%). Few seems to be ready to analyse and take advantage of unstructured data, such as text or images (for profit 17%; nonprofit 10%).

	Type of data constantly analysed	
	For profit (n = 115)	Nonprofit (n = 63)
1. Data about people's online behaviours *(click-streams; logs; search histories)*	30%	27%
2. Data about individuals' economic transactions *(credit cards operations)*	30%	16%
3. Data about people's attitudes *(survey opinions)*	30%	37%
4. Data about geographical location *(GPS or mobile telephone signals)*	21%	29%
5. Data about people's attributes *(ethnicity; occupation; health conditions)*	18%	49%
6. Unstructured data like voice, text or images *(blogs; tweets; footages; videos)*	17%	10%

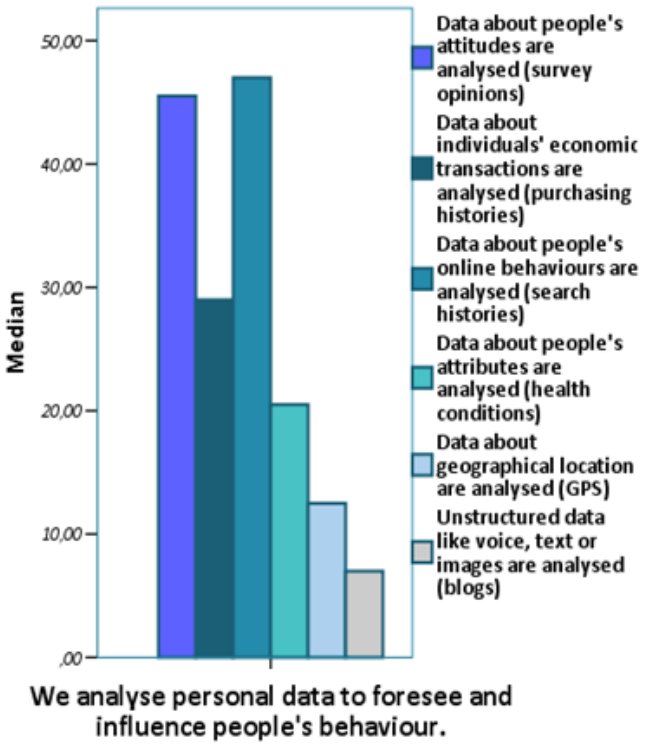
Personal Data

An important portion of all data daily processed by organisations refers to people. The fact that data are somewhat related to real persons changes dramatically the way the information has to be treated. Personally Identifiable Information (PII) is information that could potentially identify a specific individual alone or in conjunction with other types of information. "Personal data", which refers to "information relating to an identified or identifiable natural person" [6], is the expression commonly used in Europe to refer to PII. The protection of personal data is recognised as a fundamental human right since 2009, with the entry into force of the Treaty of Lisbon. Organisations operating in the European Union must comply with national laws enacting the provisions contained in the Data Protection Directive 1995.

From data gathering to data analysis and strategic action, each step of the data value chain distributes costs and benefits to different actors. Some firms can make money by selling data (8%; n = 369), while others may generate income by storing (11%) or analysing data (18%) for other organisations. Sometimes data are analysed in search of answers about human behaviour.



Information about people's characteristics and activities can be used to foresee and influence individual behaviour. The graph on the right, displays the value of the median - which is the number separating the higher half of a data sample from the lower half - for each type of data analysed, measured on a scale from 0 = "Type of data not analysed" to 100 = "Type of data constantly analysed", by organisations which try to foresee people's behaviour.



Organisations may try to influence people's behaviour by changing website contents and appearance: social media and other internet-based platforms have become huge experimental laboratories [7]. Another important source of information for understanding people's perceptions comes from public opinion pool surveys. Purchasing histories, personal characteristics, geospatial and unstructured data can also be of some utility.

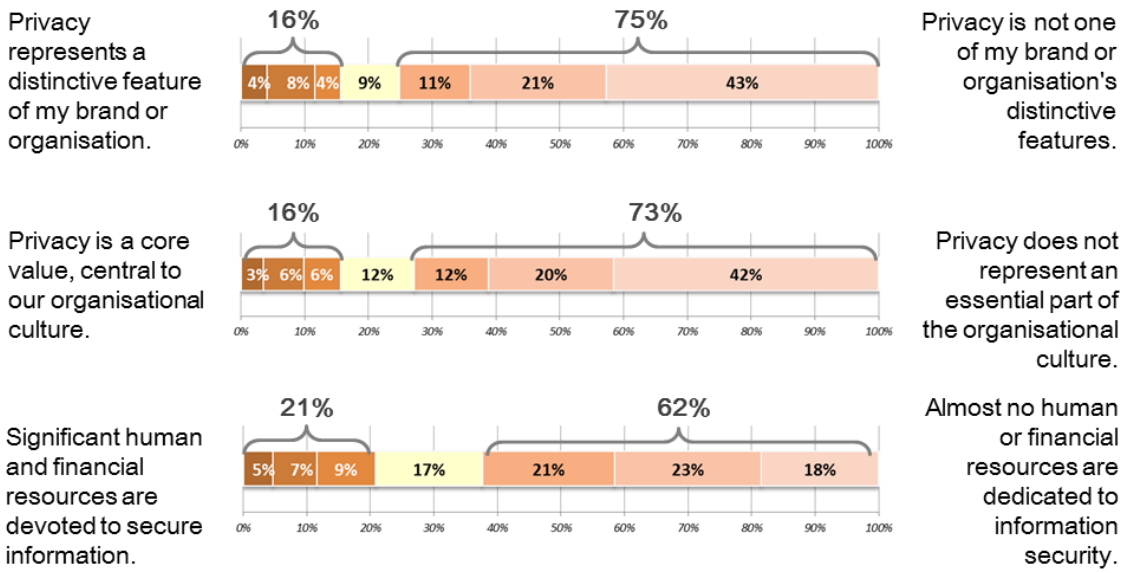
Organisational Privacy Culture

Several studies investigating people's privacy expectations demonstrate that individuals have the tendency to state how important privacy is to them and then engage in data sharing and other risky practices, which contradict their assertion [8]. People's information privacy concerns differ across cultures [9], and privacy expectations of consumers and business representatives may vary considerably [10].

Protecting the privacy of personal information poses significant challenges for those organisations which want to merely comply with laws and regulations rather than to create a culture of ethical integrity [11].

A few participants (16%; n = 173) were able to say that privacy, and its safeguard, was considered a central element of their organisational cultures and a distinctive organisational feature (16%). Only a limited number of organisations seem also to devote considerable human and financial resources to secure information (21%). Accordingly, when asked what measures had actually been adopted, few respondents could demonstrate that their organisations embrace a comprehensive data protection approach.

Concerning your organisation's approach to individuals' privacy, such as customers' privacy, and information security, could you please rate each of the following series of statements on a scale of 1 to 7, with opposing views at either end of the scale?
(N = 173)



Safeguarding Information Privacy: EU Data Protection Principles

Even though organisations process much more data than those which fall into the category of personal data, in Europe they need to establish procedures to manage PII with special care. EU *Data Protection Directive* 1995 requires data controllers, that is, the organisations processing personal data, to respect a few basic principles. EU Data Protection Principles (DPPs) can be divided into two main categories: Data Subjects' Rights and Data Controllers' Obligations. DPPs have a lot in common with the United States Federal Trade Commission's Fair Information Practice Principles (FIPPs) and OECD *Guidelines on the Protection of Privacy and Transborder Flows of Personal Data* 1980.

Data Controllers' Obligations as stated in the Data Protection Directive 1995 and in the UK Data Protection Act 1998:

- **Data Accuracy**: Individuals' data should be kept complete, accurate and up-to-date.
- **Data Minimisation**: Only the minimum amount of personal data necessary to fulfil a specific objective should be collected.
- **Data Sharing**: Individuals' data should be shared only with authorised third parties.
- **Data Retention**: Personal data should be deleted once the objective for which they have been collected is achieved.
- **Data Security**: Strong security measures should be adopted to protect data from unauthorised use.
- **Accountability**: Procedures to compensate individuals in case data were lost, manipulated or stolen, as well as sanction for those who use or handle personal data inappropriately, should be put in place.

Data Subjects' Rights, as envisioned in the EU Data Protection Directive 1995 and in the UK Data Protection Act 1998:

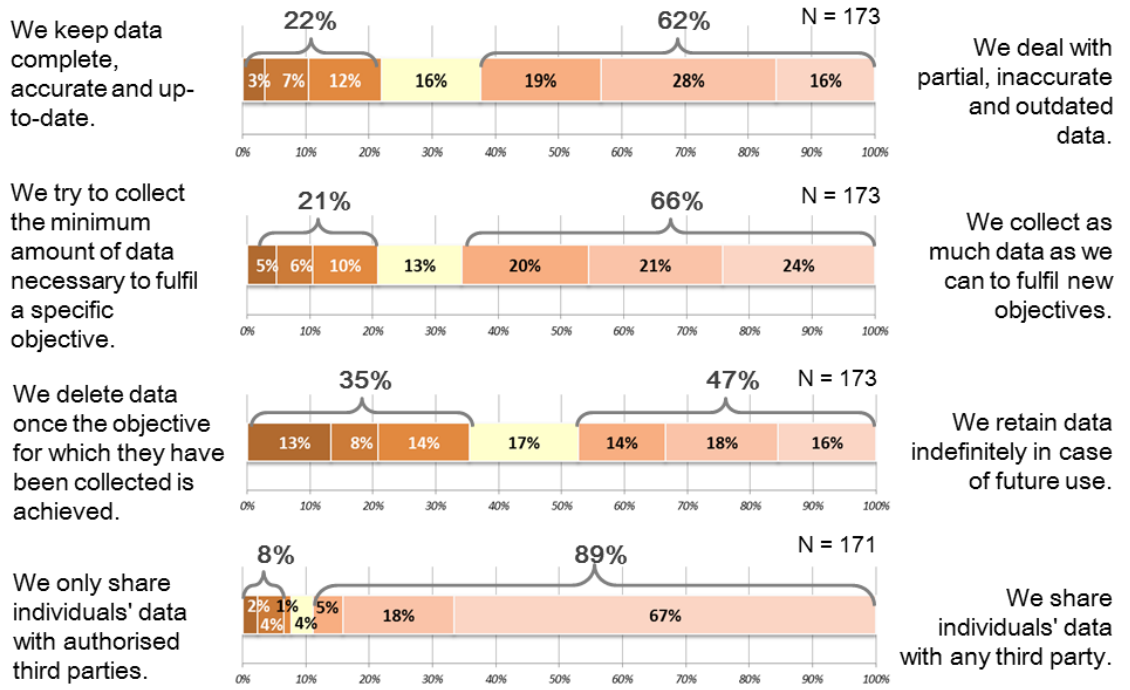
- **Right of Notice**: Individuals are informed about all aspects related to the processing of their data.
- **Right to Consent**: Individuals are asked to give their explicit consent to the processing of their data.
- **Right of Access**: The organisation has procedures in place to let the individuals access and rectify inaccurate data.
- **Right to Object**: The organisation has procedures in place to satisfy individuals' requests to end the processing of their data.

Data Controllers' Obligations

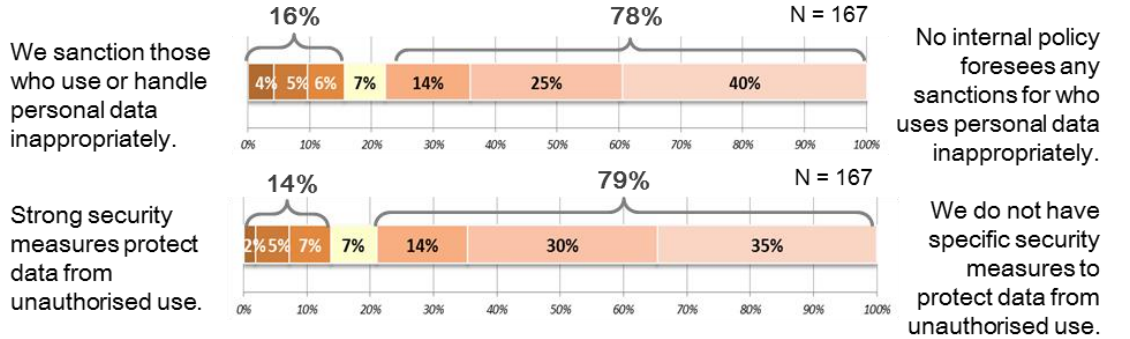
Ensuring good data quality seems to be something very difficult to achieve. 62% of respondents admitted they deal with partial, inaccurate and outdated data. Nonetheless, most organisations keep collecting as much data as they can to fulfil new objectives (66%), while almost half of study participants said their organisations retain data indefinitely and potentially for future use (47%; n =

173). Moreover, 67% of respondents said their organisations share individuals' data with any third party.

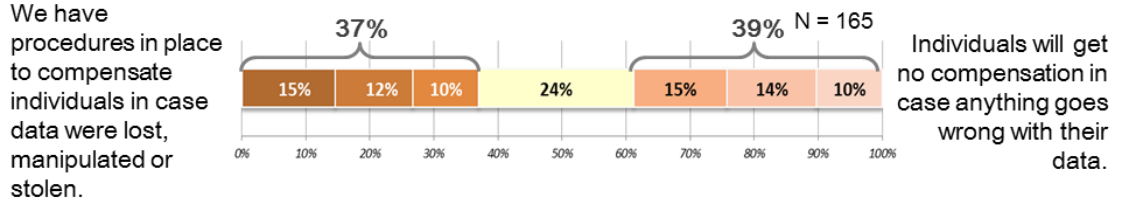
Please indicate which statement better reflects your organisation's approach to the management of individuals' data, on a scale of 1 to 7, with opposing views at either end of the scale.



Organisations seem also not to foresee any sanctions for those employees who use personal data inappropriately (40%). Regarding information security, 35% of respondents said their organisations do not have specific security measures to protect data from unauthorised use (n = 167). Yet in some cases individuals do get some compensation in the event of something going wrong with their data (37%; n = 165).



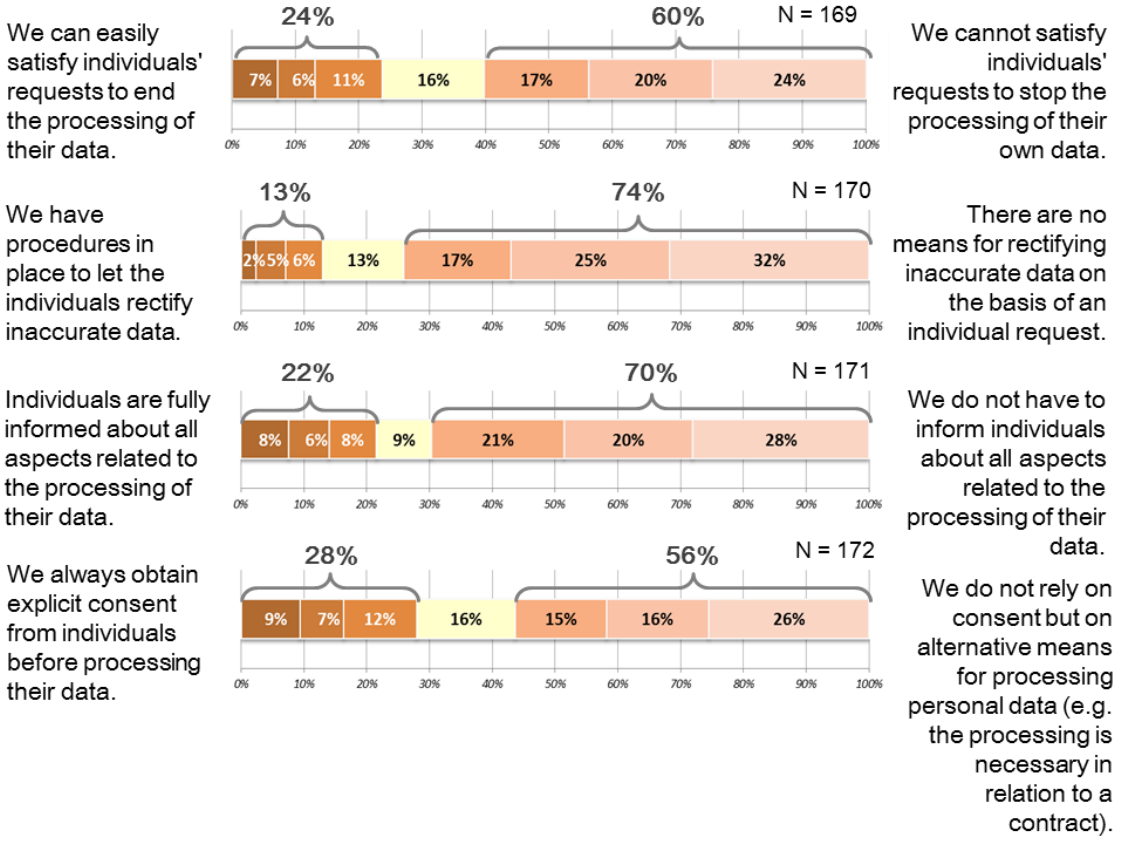
Please indicate which statement better reflects your organisation's approach to the management of individuals' data, on a scale of 1 to 7, with opposing views at either end of the scale.



Data Subjects' Rights

Protecting data subjects' rights seems to be especially problematic: 60% of participants said that their organisations cannot satisfy individuals' requests to stop the processing of their personal data, while 74% said that there are no means for rectifying inaccurate data on the basis of an individual request [12]. In addition, 70% of respondents considered that their organisations do not have to inform individuals about all aspects related to the processing of their data. This might also be due to the fact that the majority of organisations do not rely on consent, but on alternative means, such as the contract terms and conditions, for processing personal data (56%; n = 172).

Please indicate which statement better reflects the way your organisation handles personal data, on a scale of 1 to 7, with opposing views at either end of the scale.



Data Protection Regulatory Regime

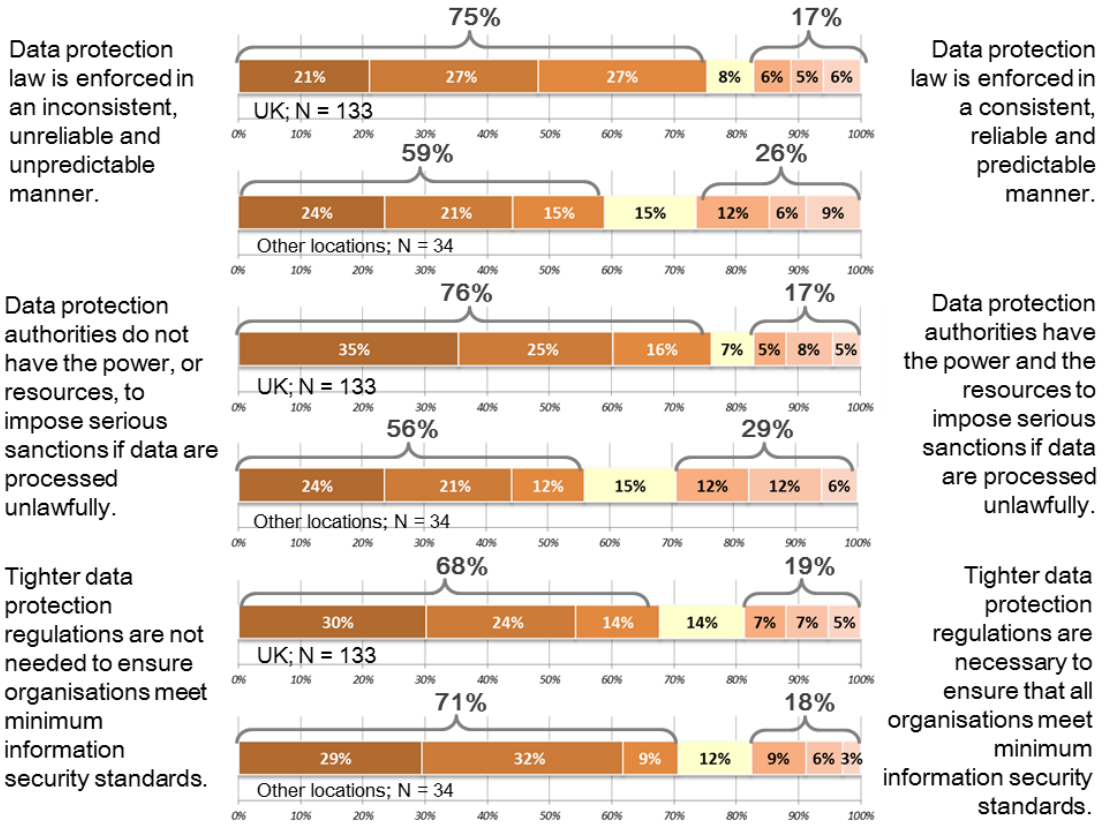
As an important part of complying with data protection principles depends on the enactment of the law and on the powers of national Data Protection Authorities (DPAs), we asked study participants to answer a few questions regarding their perceptions of their national data protection regimes. Since the large majority of organisations, which participated in the study, were based in the UK, responses of British participants are used as a benchmark.

Although three quarters of British participants said that data protection law in the United Kingdom is not enforced in a reliable manner [*], and that the national DPA, that is, the Information Commissioner's Office (ICO), in their opinion, does not have the power, or the resources, to impose serious sanctions in cases where

data are processed unlawfully, few participants considered that tighter legislation would be necessary to ensure minimum information security standards are met by organisations (19%; n = 133).

The majority of participants based in other locations considered the way data protection law is enforced in their countries equally problematic (n = 34). 59% of respondents said that data protection law is enforced in an unpredictable manner, while 56% said that data protection authorities cannot impose serious sanctions. However, having a more robust data protection regime was considered a good solution only by very few respondents (18%; n = 34).

Concerning data protection regulation in your country, could you please rate each of the following series of statements on a scale of 1 to 7, with opposing views at either end of the scale?



Proposed General Data Protection Regulation

On the 25th of January 2012, the European Commission proposed a comprehensive reform of the EU Data Protection Directive 1995 with the aim of harmonising data protection laws across EU member states. The legal instrument chosen was not a directive but a Regulation, immediately applicable across EU member states. In March 2014 the European Parliament approved an amended version of the new proposed General Data Protection Regulation (GDPR), which is currently at the centre of negotiations between the European Parliament, the Council of Ministers and the European Commission.

Now three years since the publication of the first draft regulation, 45% of respondents said their organisation have already begun planning for the new Regulation (n = 167). With respect to the provisions contained in the text approved by the European Parliament last year, the right to erasure (72%) and the right to data

portability (66%) were considered somewhat or highly problematic by the large majority of professionals. Organisations making money by selling, analysing, or storing data find particularly problematic the provisions on data portability (66%) and explicit consent (42%; n = 53) instead.

Percentage of respondents which consider each provision somewhat or highly problematic		GDPR Provision
All organisations (n = 167)	Organisations selling, analysing, and storing data* (n = 53)	
72%	68%	Data subjects will have the right to erasure. This will allow individuals to have all personal data that business holds on them deleted or restricted.
66%	66%	Data subjects will have the right to data portability, which is a right to require a portable copy of a data subject's personal data so that they may transfer it to another data controller.
60%	55%	The regulation will apply to organisations outside the EU whenever they process personal data of individuals in the EU. Data transfer outside the EU will be possible through Binding Corporate Rules (BCR) or in case of authorisation given by data protection authorities. Authorisations will be valid only for two years.
53%	47%	Privacy Impact Assessment (PIA) will have to be performed annually, and organisations will be encouraged to adopt Privacy-by-Design principles (PbD) and to certify their data processing procedures by a supervisory authority and accredited third-party auditors.
38%	42%	Consent must be given by a data subject in a clear statement or via an affirmative action (i.e. ticking a consent box when visiting a website) in cases when explicit consent would be required.
35%	43%	Serious data breaches will have to be notified to both the Data Protection Agency and data subjects. Supervisory authorities will maintain a public register of the types of breaches notified. Notification must be given without undue delay.
21%	23%	A data protection officer (DPO) will have to be appointed by public authorities and businesses if data of more than 5,000 data subjects is processed in any consecutive 12-month period, and if (i) special categories of data, (ii) location data, (iii) data relating to children, or (iv) employee data in large scale filing systems, are processed.

* These organisations come from the technology (31%) and the telecommunication (12%) sectors, as well as from professional (27%) and financial service (13%) industries. Marketing-intense sectors, such as the media and entertainment (4%) or in the hospitality (2%) sectors, were also represented in this group.

Specific Privacy and Security Safeguards

When considering the kinds of concrete measures organisations implement to protect data, it is worth noting that many organisations have detailed data privacy policies (80%; n = 167), and that a large

number of organisations employ a Chief Privacy Officer (65%). In most cases, employees receive privacy training (60%), and can be sanctioned in case of serious data mismanagement (53%). Opt-in procedures (49%) are more commonly applied than opt-out procedures (37%) at the time of asking individuals to consent to the processing of their personal data. Privacy-by-Design principles are rarely implemented (28%).

Privacy Impact Assessments (32%) or Privacy Enhancing Technologies (PETs) (27%) are also infrequently used.

Regarding information security, encryption is commonly used to protect data both on discs (55%) and during transmission (59%). External auditors' assessments (44%) and vulnerability tests (59%) are also undertaken by some organisations to evaluate their degree of resilience. Few organisations have security certifications (35%) or procedures in place to notify individuals in case of a data breach (26%). Even less invest in data breach insurance policies (16%) or rely on Binding Corporate Rules (BCRs) to manage international data transfers (15%).

No	Measures and procedures adopted (n = 167)	Freq.	%	No	Measures and procedures adopted (n = 167)	Freq.	%
1.	Data policies, which describe the rules controlling the integrity, security, quality, and use of data during its life-cycle and state change, have been adopted.	133	80%	2.	A Chief Privacy/Data Protection Officer (DPO) is in charge of supervising all privacy-related issues.	109	65%
3.	The function of dealing with privacy-related matters is pursued by a designated department inside my organisation, for example the compliance office or the IT department.	105	63%	4.	Employees are constantly trained to comply with privacy procedures.	100	60%
5.	Specific policies for classifying information according to their sensitivity (secret; confidential; etc.) are in place.	99	59%	6.	Network and application penetration and vulnerability testing ('friendly hacking').	99	59%
7.	Encrypted transmission of data.	98	59%	8.	Full-disk encryption of physical devices like laptops or PCs.	92	55%
9.	Workforce members are sanctioned if they do not comply with privacy procedures.	89	53%	10.	Consent obtained through opt-in acceptance of data processing terms and conditions.	82	49%
11.	Periodical external auditors' assessment of internal security standards.	73	44%	12.	Consent obtained through opt-out acceptance of data processing terms and conditions.	61	37%
13.	Certified code of practice for information security management (ISO/IEC 27002:2005).	59	35%	14.	Counsel of a legal firm specialized in information privacy.	55	33%
15.	Privacy Impact Assessments (PIAs) are undertaken.	54	32%	16.	Privacy-by-design (PbD) criteria are adopted in product development.	47	28%
17.	Privacy Enhancing Technologies (PETs) are in use.	45	27%	18.	Immediate notification to individuals if their data are disclosed or manipulated.	44	26%
19.	Data breach insurance policy.	26	16%	20.	Binding Corporate Rules (BCRs) to manage international data transfer.	25	15%

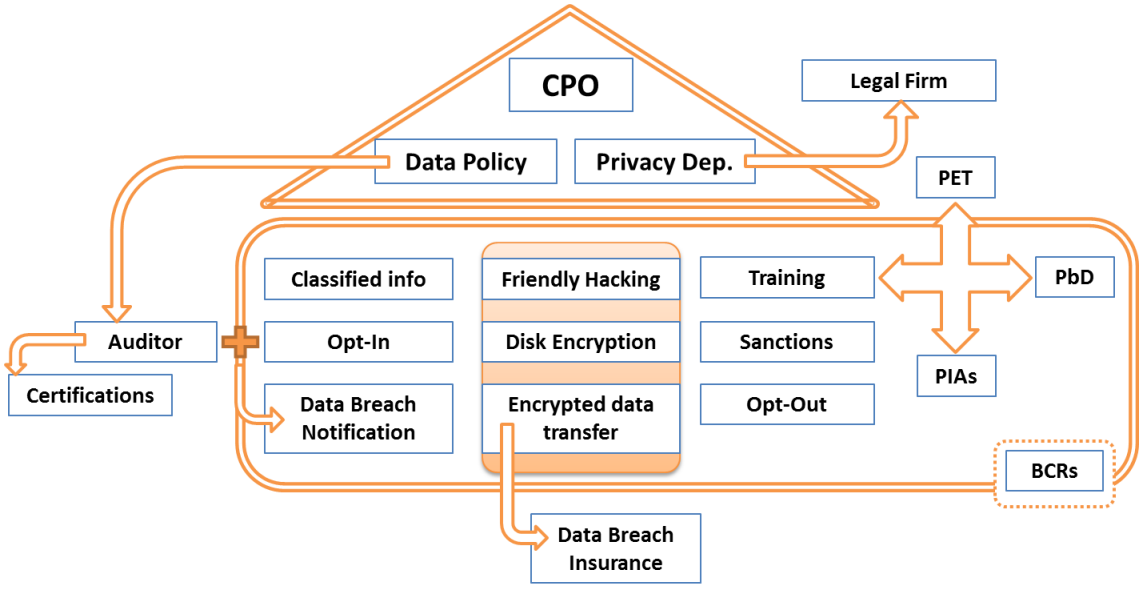
Relationships between Various Privacy and Security Safeguards

The diagram below shows which privacy and security measures are typically implemented together [13]. The presence of a Data Privacy Officer within a specialised internal department fosters the adoption of detailed data handling policies. Some organisations rely on the advice and services of an external consultant or legal firm, or foresee the appointment of an external auditor,

to monitor the implementation of their data management policies.

The internal privacy team works in two fundamental areas: (a) the privacy training of the workforce; and (b) the resilience of the security system. The demand for privacy enhancing technology is often driven by workforce training needs. Initiatives such as running privacy impact assessments, or adopting privacy by design principles, are managed as part of special training activities.

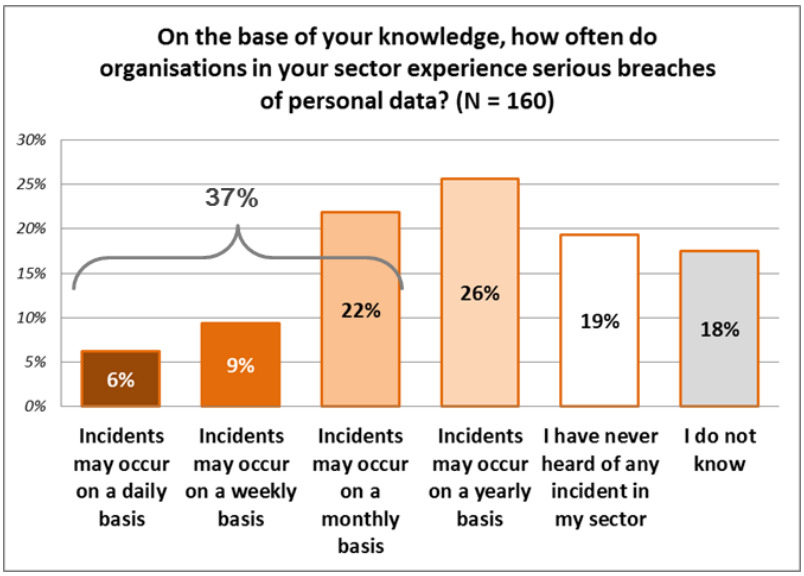
Cybersecurity measures, such as encryption or friendly hacking, are usually part of the same information security strategy. Organisations which have data breach insurance policies also use encrypted data transfer. Organisations which adopt opt-in procedures and appoint an external auditor are more likely to enact data breach notification policies.



The Rationale behind investing in Information Security

According to the results of this study, organisations can take either a proactive or a reactive approach toward investing in information security [14]. Reactive organisations invest in information security as a response to previous security problems: their investments in information security are mostly motivated by risks of high economic loss, costly enforcement action by regulators or the payment of high litigation costs. In contrast, proactive organisations are more concerned about potential reputational risks of a major security incident, meeting high industry information security standards, or improving the quality of their services or products. There are differences between for profit and nonprofit organisations in this respect. Public institutions are especially concerned about being sanctioned by regulators (62%), while private companies worry more about serious economic losses (63%) and high industry standards (63%). Nonetheless, all organisations are concerned about their reputation (for profit 77%; nonprofit 71%). Finally, suffering a data breach represents a monthly or more frequent type of event for 37% of respondents (n = 160).

Reasons for investing in information security				
	For profit (n = 115)		Nonprofit (n = 63)	
	Freq.	%	Freq.	%
1. To manage reputational risks	88	77%	45	71%
2. To reflect high industry security standards	73	63%	36	57%
3. To manage the risk of economic loss	73	63%	32	51%
4. To avoid costly enforcement action by regulators	67	58%	39	62%
5. To improve service/product quality	66	57%	31	49%
6. To manage the risk of high litigation costs	62	54%	25	40%
7. To react to previous security problems	51	44%	26	41%



Explaining Data Protection Practices

As a very large portion of all the data processed by organisations refer - directly or indirectly - to real persons, the question of the relationship between big data and information privacy becomes a central one. It is very controversial as to what kinds of incentives - or external conditions - may motivate organisations to secure data and respect people's privacy rights. The study asks a simple question: is it possible to create synergies between big data innovation and the respect of data protection principles? How? Some commentators argue that, since analytically sophisticated organisations value data, they are more likely to invest in information security and to adopt well-defined privacy policies about customer and employee information [15]. Can, thus, enterprises competing on analytics be privacy champions?

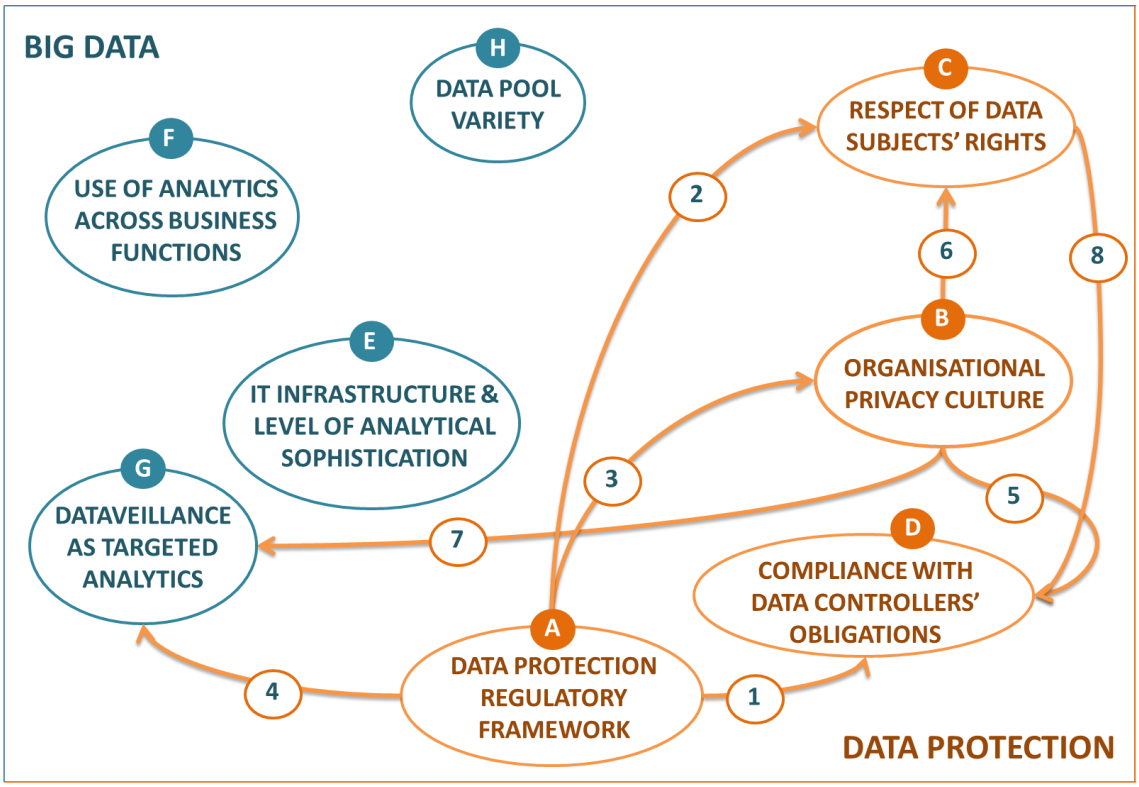
In trying to shed light on the drivers behind the implementation of fair information management procedures, and to unveil the complex set of relationships between big data analytics and data protection, three main elements have been investigated. These factors are: the degree of clarity and permissiveness of the data protection regulatory framework in force in the country where the organisation operates; the organisation's internal privacy culture; the level of IT and analytical sophistication an organisation has achieved.

Because of the complexity and large number of dimensions analysed, first we focus on those aspects more strictly related to data privacy, then to big data, and finally we combine both realms into the same picture. The next chart shows the relationships between four key dimensions related to privacy and data protection: (A) the data protection regulatory regime; (B) the organisation's internal privacy culture; (C) the respect of data subjects' rights; and (D) the level of compliance with data controllers' obligations. The second chart explores the relationships between another four dimensions, this time related to big data. These are: (E) the level of IT integration and analytical sophistication; (F) the use of analytics across business functions; (G) the reliance on dataveillance and targeted analytics; (H) the variety of the big data pool analysed by the organisation. The hypotheses represented graphically in the next three charts, have been tested by means of a statistical technique called structural equation modelling, with generalised least squares estimator (n = 195). All hypotheses here presented found support in the analysis. For further information on model fit and other parameters the reader is invited to contact the author. Each hypothesis is identified by a number on an arrow linking two concepts. Lines in dashes indicate negative relationships, while full lines indicate positive relationships.

Relationships between Data Protection Dimensions

Regulation plays a very important role in shaping the organisational environment and in influencing organisational choices. (A) Countries where data protection law is perceived to be enforced in a consistent, reliable and predictable manner, and data protection authorities (DPAs) have the power, and the resources, to impose serious sanctions, feature organisations (1) which comply with data controllers' obligations, (2) which respect data subjects' rights, (3) consider privacy a core value, central to the organisational culture, and (4) which use analytics to achieve specific objectives.

Despite the importance of the regulatory approach, privacy needs to be part of the organisational culture, and to be valued by both leaders and employees, in order to transform organisational practices. (B) Organisations which consider privacy a core value, and devote considerable human and financial resources to secure information, are in fact typically (5) compliant with data controllers' obligations, (6) respectful of data subjects' rights, but also (7) capable of applying analytics in a targeted way. In addition, (C) organisations which respect data subjects' rights, are (8) typically fully compliant with (D) data controllers' obligations.



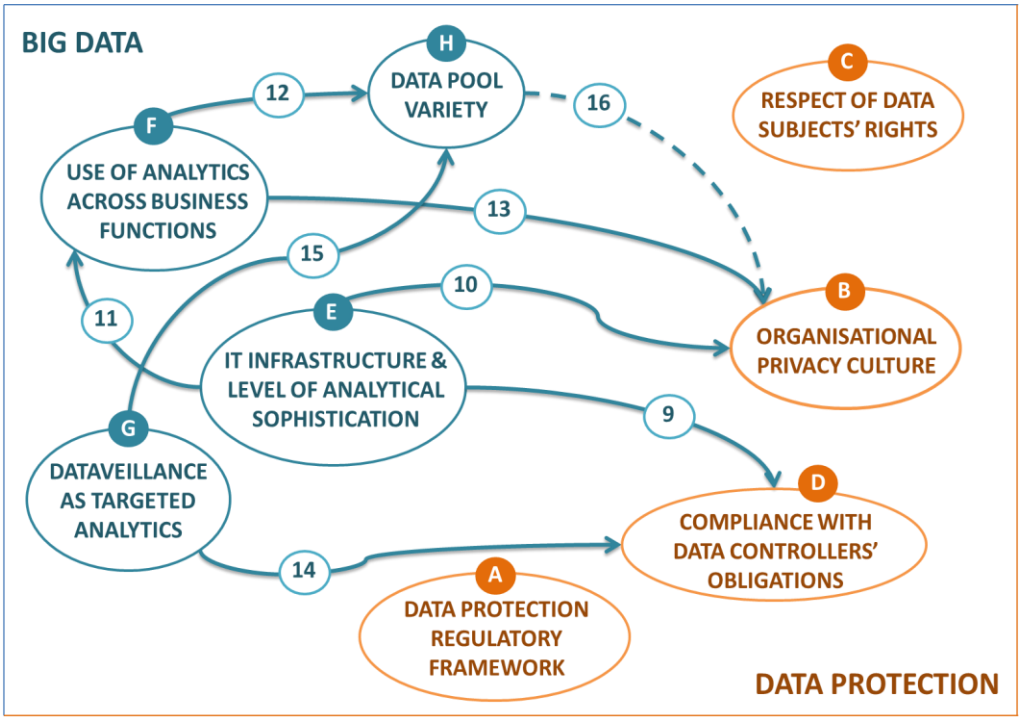
Relationships between Big Data Dimensions

Investing in the organisation's information management system is a precondition to both protect data and become analytically sophisticated. In fact, (E) organisations which rely on analytics and have a flexible, integrated IT infrastructure to work with data, tend to (9) be more compliant with data controllers' obligations, (10) consider privacy central to the organisational culture, and (11) use analytics across business functions.

(G) Organisations which collect data to monitor and influence individuals' activities, and which rely on profiling to target valuable users, or personalise offers, tend to (14) comply with data controllers' obligations, despite (15) they collect and process a diverse array of data.

(F) Organisations which use analytics across business functions, that is, to foster marketing, improve security, gain efficiency, to better manage human resources, to reduce financial risks, and to take better informed strategic decisions, despite being the ones (12) which collect more data of different kinds, (13) tend to consider privacy a core value, central to the organisational culture. Nonetheless, the logic of big data and the logic of data protection collide on data collection.

(H) Organisations which manage large amounts of data of different kinds - from geographical locations, to people's online behaviours, economic transactions, attributes and attitudes -, are less willing or capable - to (16) transform privacy into a core value, central to the organisational culture. Although data harvesting and privacy seem to remain fundamentally incompatible, new opportunities to create privacy-enhancing technologies, as well as sophisticated data minimisation and de-identification tools may emerge from the same digital innovation environment fostering the big data revolution [16].



CONCLUSION: Analytical Sophistication and Data Privacy - the Full Picture

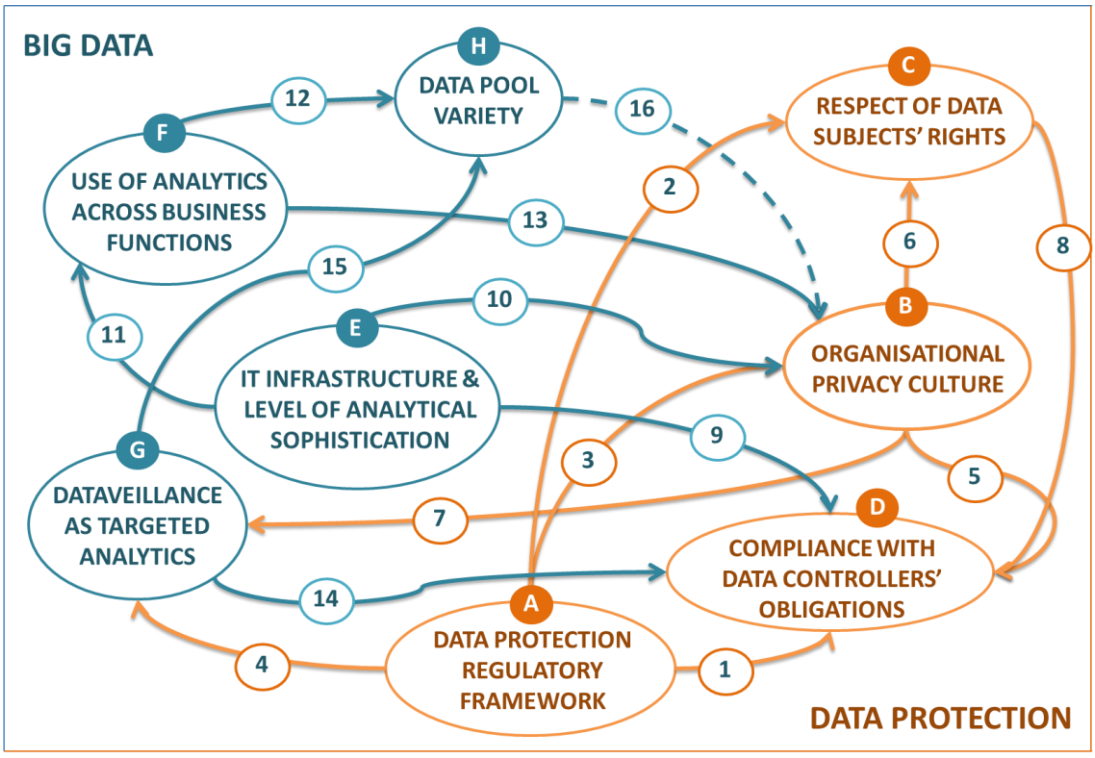
A strong and reliable data protection regime represents the precondition to foster high data privacy standards across all type of organisations, whether private or

public. Regulation is especially important for ensuring the respect of data subjects' rights and to embed privacy into the organisational culture. Although institutional pressure is beneficial overall, the way privacy is embedded into the organisational culture is even more important as it shapes the way the organisation uses analytics to achieve its targets.

Investing in the organisation's information management infrastructure is absolutely important both for privacy and competitiveness. The more data becomes a central strategic element for the organisation, the more privacy and the privacy function have to engage with all other analytically-driven business functions.

The uncontrolled collection of data of any kind is the only element in clear opposition with fostering an authentic privacy culture within an organisation. The massive gathering of data, in different formats stored for future use, seems to be the real point of friction between the big data and the data protection worlds.

Synergies between data protection and big data can be found on the terrain of data quality and targeted analytics, but not on data harvesting as a goal by itself. The big data revolution is misleadingly said to be driven just by data accumulation: analytics, the real value-added component of big data, seems not only to be compatible with data protection norms, but even to benefit from a strong organisational privacy culture and a clear data protection regulatory regime.



Notes

[1] ICO e-newsletter, December 2013, Your Thoughts: Join the Open University's 'Big Data Protection' Study", available at: http://ico.msgfocus.com/q/1bkJRNvP4UxkK503bL5/wv . ICO e-newsletter, February 2014, "Last chance to join the Open University's 'Big Data Protection' Study". Available at: http://ico.msgfocus.com/q/1AFxOgho8z/wv

[2] The Operation Research Society's magazine, available at: http://www.theorsociety.com/

[3] More information available on the ELITE Group's website: http://www.bcs.org/category/18242

[4] Author's elaboration of the DELTA model presented in Davenport, Thomas H., Jeanne G. Harris, and Robert Morison, 2010, 'Analytics at Work: Smarter Decisions, Better Results', Boston (MA): Harvard Business Press, p. 19.

[5] Degli Esposti, Sara, 2014, "When big data meets dataveillance: The hidden side of analytics." Surveillance & Society, 12(2): 209-225, available at: http://library.queensu.ca/ojs/index.php/surveillance-and-society/article/view/analytics

[6] "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data", Official Journal L 281 , 23/11/1995 P. 0031 - 0050, available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML

[7] "Arthur, Charles, 2014, "Facebook emotion study breached ethical guidelines, researchers say", The Guardian, Monday 30 June 2014 09.51 BST, available at: http://www.theguardian.com/technology/2014/jun/30/facebook-emotion-study-breached-ethical-guidelines-researchers-say

[8] See, for example, Norberg, Patricia A., Daniel R. Horne, and David A. Horne, 2007, "The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors ", Journal of Consumer Affairs, 41(1): 100-126, doi: 10.1111/j.1745-6606.2006.00070.x.

[9] Milberg, Sandra J., Sandra J. Burke, H. Jeff Smith, and A. Kallman Ernest, 1995, "Values, Personal Information Privacy, and Regulatory Approaches ", Communications of the ACM, 38(12): 65-74.

[10] Milne, George R., and Shalini Bahl, 2010, "Are There Differences Between Consumers' and Marketers' Privacy Expectations? A Segment- and Technology-Level Analysis ", Journal of Public Policy & Marketing 29(1): 138-149.

[11] Culnan, Mary J., and Cynthia Clark Williams, 2009, "How Ethics Can Enhance Organizational Privacy: Lessons From The ChoicePoint and TJX Breaches ", MIS Quarterly 33(4): 673-687.

[12] "IRISS study finds 4 in 10 organisations obstruct our access to our own data", article posted by Hayley Watson on June 23, 2014, available at: http://irissproject.eu/?p=526

[13] The strength of association between the implementation of each measure has been measured by using the mean square contingency coefficient, also known as Phi coefficient. Please contact the author if you want to obtain the full table of coefficients for each pair of variables.

[14] This statement is based on the result of an Exploratory Factor Analysis (EFA) run on all variables presented in this section; number of factors extracted two, without rotation. Please contact the author for more information.

[15] Davenport, Thomas H., Jeanne G. Harris, and Robert Morison, 2010, *Analytics at Work: Smarter Decisions, Better Results*, Boston (MA): Harvard Business Press. On page 34 is written: "Stage 5 firms [Analytical Competitors] follow the Hippocratic oath of information privacy: above all, they do not harm. They have well-defined privacy policies… They don't break the privacy laws… They don't lose information… They don't sell or give away information without the permission of the customer or employee."

[16] Cavoukian, Ann, and Daniel Castro, 2014, "Big Data and Innovation, Setting the Record Straight: De-identification Does Work", available from: https://www.privacybydesign.ca/content/uploads/2014/06/pbd-de-identification_ITIF1.pdf

*** Disclaimer**: Percentages included in this report are indicative only and should be treated with caution; because of the non-probability sampling design adopted, opinions here expressed should not be considered representative of the opinions of the entire population of British data controllers. For a more accurate view on stakeholders' perceptions of the ICO's activities, please see the report "Stakeholder Perceptions 2012" prepared by Ipsos MORI in 2012 for the Information Commissioner's Office", and available at: https://ico.org.uk/media/about-the-ico/documents/1042371/stakeholder-perceptions-2012.pdf

References

Aaronson, S. A. and R. Maxim (2013). "Data Protection and Digital Trade in the Wake of the NSA Revelations." Intereconomics **48**(5): 281-285.

Abine. (2015). "DoNotTrackMe: How it Works." from https://www.abine.com/donottrackme.html.

Acquisti, A. (2010). Background Paper #3: The Economics of Personal Data and the Economics of Privacy. The Economics of Personal Data and Privacy: 30 Years after the OECD Privacy Guidelines, OECD Conference Centre.

Acquisti, A. (2010). Background Paper #3: The Economics of Personal Data and the Economics of Privacy. The Economics of Personal Data and Privacy: 30 Years after the OECD Privacy Guidelines, OECD Conference Centre, OECD.

Acquisti, A., L. Brandimarte and G. Loewenstein (2015). "Privacy and human behavior in the age of information." Science 509-514.

Acquisti, A., A. Friedman and R. Telang (2006). Is there a cost to privacy breaches? An event study. International Conference of Information Systems (ICIS), Milwaukee (WI), Association for Information Systems.

Acquisti, A. and R. Gross (2009). "Predicting Social Security numbers from public data." Proceedings of the National Academy of Sciences of the United States of America **106**(27): 10975-10980.

Acquisti, A. and H. R. Varian (2005). "Conditioning Prices on Purchase History." Marketing Science **24**(3): 367-381.

Adams, N. M. (2010). "Perspectives on data mining." International Journal of Market Research **52**(1): 11-19.

Agresti, A. (2010). Analysis of Ordinal Categorical Data. Hoboken, New Jersey, Wiley.

Agresti, A. (2013). Categorical Data Analysis. Hoboken: New Jersey, Wiley-Interscience.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. Proceedings of the 2nd International Symposium on Information Theory, Budapest: Akademiai Kiado.

Akaike, H. (1987). "Factor analysis and AIC." Psychometrika **52**: 317–332.

Altman, I. (1975). The Environment and Social Behavior: Privacy, Personal Space, Territory, and Crowding. Monterey, California 93940, Brooks/Cole Publishing Company.

Amoore, L. and M. DeGoede (2005). "Governance, risk and dataveillance in the war on terror." Crime, Law and Social Change **43**(2): 149-173.

Anderson, B. and J. M. Hardin (2014). Streaming Data in the Age of Big Data. Big Data, Mining, and Analytics. S. Kudyba, Auerbach Publications**:** 165-178.

Anderson, C. (2009). Chris Anderson interview. The Virtual Revolution:The Cost of Free. D. Biddle, BBC TWO.

Anderson, R. (2001). Why Information Security is Hard-An Economic Perspective. ACSAC '01 Proceedings of the 17th Annual Computer Security Applications Conference, IEEE Computer Society Washington, DC, USA.

Andrejevic, M. (2009). "Control Over Personal Information in the Database Era." Surveillance & Society **6**(3): 322-326.

Andrejevic, M. (2009). iSpy: Surveillance and Power in the Interactive Era, University of Kansas Lawrence, KS, USA.

Antón, A. I., J. B. Earp and R. A. Carter (2003). "Precluding incongruous behavior by aligning software requirements with security and privacy policies." Information and Software Technology **45**(14): 967-977.

Art29 (2009). Thirteenth Annual Report on the situation regarding the protection of individuals with regard to the processing of personal data and privacy in the European Union and in third countries, Article 29 Data Protection Working Party.

Art29 (2010). European Data Protection Authorities meet with European Commission Vice-President Viviane Reding to discuss the review of the data protection framework. ARTICLE 29 DATA PROTECTION WORKING PARTY. Brussels.

Art29 (2010a). European Data Protection Authorities meet with European Commission Vice-President Viviane Reding to discuss the review of the data protection framework. Brussels, Article 29 Data Protection Working Party.
Art29 (2010b). Report 01/2010 on the second joint enforcement action: Compliance at national level of Telecom Providers and ISPs with the obligations required from national traffic data retention legislation on the legal basis of articles 6 and 9 of the e-Privacy Directive 2002/58/EC and the Data Retention Directive 2006/24/EC amending the e-Privacy Directive, Article 29 Data Protection Working Party. **172**.
Art29 (2014). Statement on Statement of the WP29 on the impact of the development of big data on the protection of individuals with regard to the processing of their personal data in the EU Adopted on 16 September 2014, Article 29 Data Protection Working Party. **WP 221**.
Articles29 (2013). Opinion 03/2013 on purpose limitation. Adopted on 2 April 2013. A. D. P. W. Party. Brussels, Directorate C (Fundamental Rights and Union Citizenship) of the European Commission. **WP 203**.
Ashworth, L. and C. Free (2006). "Marketing Dataveillance and Digital Privacy: Using Theories of Justice to Understand Consumers' Online Privacy Concerns." Journal of Business Ethics **67**(2): 107-123.
Askin, F. (1972). "Surveillance: The Social Science Perspective." Columbia Human Rights Law Review **4**: 59-88.
Awad, N. F. and M. S. Krishnan (2006). The Personalization Privacy Paradox: An Empirical Evaluation Of nformation Transparency And The Willingness To Be Profiled Online For Personalization. MIS Quarterly, MIS Quarterly & The Society for Information Management. **30:** 13-28.
Ayres, I. (2007). Super crunchers: Why thinking-by-numbers is the new way to be smart, Bantam Books.
Babin, B. J., J. F. Hair, Jr. and J. S. Boles (2008). "Publishing research in marketing journals using structural equation modeling." Journal of Marketing Theory and Practice **16**(4): 279-285.
Bagozzi, R. P., Ed. (1994). Principles of Marketing Research. Cambridge, Mass, Blackwell Business.
Bagozzi, R. P. (1994). Structural Equation Models in Marketing Research: Basic Principles. Principles of Marketing Research. R. P. Bagozzi. Cambridge, UK, Blackwell Publishers**:** 317-385.
Bagozzi, R. P., Y. Yi and L. W. Phillips (1991). "Assessing Construct Validity in Organizational Research." Administrative Science Quarterly **36**(3): 421-458.
Bajaj, K. (2012). "Promoting Data Protection Standards through Contracts: The Case of the Data Security Council of India." Review of Policy Research **29**: 131-139.
Baker, S. (2009). What Data Crunchers Did for Obama. Businessweek, Bloomberg.
Ball, K. (2010). "Data protection in the outsourced call centre: an exploratory case study." Human Resource Management Journal **20**(3): 294-310.
Ball, K. and D. C. Wilson (2000). "Power, control and computer-based performance monitoring: Repertoires, resistance and subjectivities." Organization Studies **21**(3): 539-565.
Ball, K. S. and D. Murakami Wood (2013). "Political Economies of Surveillance." Surveillance & Society **11**(1/2): 1-3.
Bamberger, K. A. (2010). "Technologies of Compliance: Risk and Regulation in a Digital Age." Texas Law Review **88**(4): 669-739.
Bamberger, K. A. and D. K. Mulligan (2011). "New Governance, Chief Privacy Officers, and the Corporate Management of Information Privacy in the United States: An Initial Inquiry." Law & Policy **33**(4): 477-508.
Bamberger, K. A. and D. K. Mulligan (2011). "Privacy on the Books and on the Ground." Stanford Law Review **63**(2): 247-315.
Bankston, K. S. and A. Soltani (2014). "Tiny Constables and the Cost of Surveillance: Making Cents Out of United States v. Jones." The Yale Law Journal Online **123**: 335-357.
Barnes, M. E. (2006). "Falling Short of the Mark: The United States Response to the European Union's Data Privacy Directive." Northwestern Journal of International Law & Business **27**(1): 171-197.
Baruch, Y. and B. C. Holtom (2008). "Survey response rate levels and trends in organizational research." Human Relations **61**(8): 1139-1160.

Baskerville, R. (1993). "Information systems security design methods: implications for information systems development." ACM Computing Surveys (CSUR) **25**(4): 375-414.
Bassi, L. (2011). "Raging Debates in HR Analytics." People and Strategy **34**(2): 14-18.
Baumer, D. L., J. B. Earp and J. C. Poindexter (2004). "Internet privacy law: a comparison between the United States and the European Union." Computers & Security **23**(5): 400-412.
Baumer, D. L., J. C. Poindexter and J. B. Earp (2004). "Meaningful and Meaningless Choices in Cyberspace." Journal of Internet Law **7**(11): 3-11.
Baumgartner, H. and C. Homburg (1996). "Applications of structural equation modeling in marketing and consumer research: A review." International Journal of Research in Marketing **13**(2): 139-161.
Behar, R. (2004). Never Heard Of Acxiom? Chances Are It's Heard Of You. How a little-known Little Rock company--the world's largest processor of consumer data--found itself at the center of a very big national security debate, Fortune.
Bélanger, F. and R. E. Crossler (2011). "Privacy in the digital age: a review of information privacy research in information systems." MIS Quarterly **35**(4): 1017-A1036.
Bell, P. C. (2015). "Sustaining an Analytics Advantage." MIT Sloan Management Review **56**(3): 21-24.
Bellman, S., E. Johnson, S. Kobrin and G. Lohse (2004). International Differences in Information Privacy Concerns: A Global Survey of Consumers. Information Society, Routledge. **20:** 313-324.
Bennett, C. J. (1992). Regulating Privacy: Data Protection and Public Policy in Europe and the United States. New York, Cornell University Press.
Bennett, C. J. (2011). "In Defense of Privacy: The Concept and the Regime." Surveillance & Society **8**(4): 485-496.
Bennett, C. J. (2011). "In Further Defence of Privacy." Surveillance & Society **8**(4): 513-516.
Bentler, P. M. (1990). "Comparative fit indexes in structural models." Psychological Bulletin **107**(2): 238-246.
Bentler, P. M. and E. Freeman (1983). "Tests for stability in linear structural equation systems." Psychometrika **48**(1): 143-145.
Bernal, P. A. (2010). "Web 2.5: The Symbiotic Web." International Review of Law, Computers & Technology **24**(1): 25-37.
Bernard, E. K. and I. Makienko (2011). "The effects of information privacy and online shopping experience in e-commerce." Academy of Marketing Studies Journal **15**: 97-112.
Bernhut, S. (2012). "Big data: The new, new thing." Ivey Business Journal **76**(4): 1-1.
Bignami, F. (2007). "Privacy and Law Enforcement in the European Union: The Data Retention Directive." Chicago Journal of International Law **8**(1): 233-255.
Bikard, A. (2011). High-Profile Data Breaches Raise Security Alerts. Compliance Week, Haymarket Media, Inc. **8:** 46-47.
BIS (2014). 2014 Information Security Breaches Survey: Technical Report. Study conducted by PwC in association with InfoSecurity Europe. London, UK, Department for Business Innovation & Skills (BIS).
Bisman, J. (2010). "Postpositivism and Accounting Research : A (Personal) Primer on Critical Realism." Australasian Accounting Business & Finance Journal **4**(4): 3-25.
Blume, P. (2012). "The inherent contradictions in data protection law." International Data Privacy Law **2**(1): 26-34.
Blume, P. (2012). "Will it be a better world? The proposed EU Data Protection Regulation." International Data Privacy Law **2**(3): 130-136.
Blume, P. (2014). "The myths pertaining to the proposed General Data Protection Regulation." International Data Privacy Law **4**(4): 269-273.
Bollen, K. and R. Lennox (1991). "Conventional wisdom on measurement: A structural equation perspective." Psychological Bulletin **110**(2): 305-314.
Bollen, K. A. (1989). Structural Equations with Latent Variables. New York, Wiley.
Bollen, K. A. (1989). Structural Equations with Latent Variables, Wiley Interscience.
Bollen, K. A. and R. A. Stine (1992). "Bootstrapping goodness-of-fit measures in structural equation models." Sociological Methods and Research **21**: 205-229.
Borland, S. (2008). MI5 warns firms over China's internet 'spying'. The Telegraph.
Boyne, G. A. (2002). "Public and private management: What's the difference?" Journal of Management Studies **39**(1): 97-122.

Brandimarte, L., A. Acquisti and G. Loewenstein (2012). "Misplaced Confidences: Privacy and the Control Paradox." Social Psychological and Personality Science: 1-8.
Breckenridge, J., J. Farquharson and R. Hendon (2014). "The role of business model analysis in the supervision of insurers." Bank of England Quarterly Bulletin **54**(1): 49–57.
Bresnahan, T. F., E. Brynjolfsson and L. M. Hitt (2002). "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence." Quarterly Journal of Economics **117**(1): 339-376.
Breur, T. (2007). "How to evaluate campaign response — The relative contribution of data mining models and marketing execution." Journal of Targeting, Measurement & Analysis for Marketing **15**(2): 103-112.
Britannica (2014). Google Inc. Encyclopaedia Britannica Online Academic Edition, Encyclopædia Britannica Inc.
Britannica (2014). Search engine. Encyclopaedia Britannica Online Academic Edition, Encyclopædia Britannica Inc.
Brock, J. K. and Y. Zhou (2005). "Organizational use of the internet — scale development and validation." Internet Research **15**(1): 67–87.
Brookman, J. (2015). "Protecting Privacy in an Era of Weakening Regulation." Harvard Law & Policy Review **9**(2): 355-374.
Brown, I. (2010). Working Paper Nr. 1: The challenges to European data protection laws and principles. Comparative Study On Different Approaches To New Privacy Challenges, In Particular In The Light Of Technological Developments, LRDP KANTOR Ltd & The Centre for Public Reform.
Brown, T. A. (2006). Other Types of CFA Models: Higher-Order Factor Analysis, Scale Reliability Evaluation, and Formative Indicators. Confirmatory Factor Analysis for Applied Research. T. A. Brown. London, The Guilford Press**:** 320-362.
Brown, W. (1910). "Some experimental results in the correlation of mental abilities." British Journal of Psychology **3**: 296–322.
Browne, M. W. (1984). "Asymptotically distribution-free methods for the analysis of covariance structures." British Journal of Mathematical and Statistical Psychology **37**(1): 62-83.
Browne, M. W. and R. Cudeck (1993). Alternative Ways of Assessing Model Fit. Testing Structural Equation Models. K. A. Bollen and J. S. Long, SAGE Publications, Inc**:** 308.
Brydon, M. and A. Gemino (2008). "You've Data Mined. Now What?" Communications of AIS **2008**(22): 603-616.
Brynjolfsson, E., J. Hammerbacher and B. Stevens (2011). "Competing through data: Three experts offer their game plans." McKinsey Quarterly(4): 36-47.
Brynjolfsson, E. and L. M. Hitt (2000). "Beyond Computation: Information Technology, Organizational Transformation and Business Performance." Journal of Economic Perspectives **14**(4): 23-48.
Brynjolfsson, E. and L. M. Hitt (2003). "Computing Productivity: Firm-Level Evidence." Review of Economics & Statistics **85**(4): 793-808.
Brynjolfsson, E., T. W. Malone, V. Gurbaxani and A. Kambil (1994). "Does Information Technology Lead to Smaller Firms?" Management Science **40**(12): 1628-1644.
Brynjolfsson, E. and A. McAfee. (2013). "Is Your Company Ready for Big Data?" Harvard Business Review Retrieved 25/05/2015, from https://hbr.org/web/2013/06/assessment/is-your-company-ready-for-big-data.
Brynjolfsson, E., A. McAfee and M. Spence (2014). "New World Order." Foreign Affairs **93**(4): 44-53.
Brynjolfsson, E., H. Yu and M. D. Smith (2003). "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers." Management Science **49**(11): 1580-1596.
BVCA (2015). Guide to Cyber Security. T. Hames and J. Rashleigh, British Private Equity & Venture Capital Association and PricewaterhouseCoopers LLP**:** 12.
Calo, R. (2013). Digital Market Manipulation. Research Paper, University of Washington School of Law.
Campbell, J., A. Goldfarb and C. Tucker (2015). "Privacy Regulation and Market Structure." Journal of Economics & Management Strategy **24**(1): 47-73.

Campbell, K., L. A. Gordon, M. P. Loeb and L. Zhou (2003). "The economic cost of publicly announced information security breaches: empirical evidence from the stock market." Journal of Computer Security **11**(3): 431-448.
Cate, F. H., P. Cullen and V. Mayer-Schönberger. (2013). "Data Protection Principles for the 21st Century: Revising the 1980 OECD Guidelines." from http://www.oii.ox.ac.uk/publications/Data_Protection_Principles_for_the_21st_Century.pdf.
Cavoukian, A. and D. Castro. (2014). "Big Data and Innovation, Setting the Record Straight: De-identification Does Work." from https://www.privacybydesign.ca/content/uploads/2014/06/pbd-de-identification_ITIF1.pdf.
Cavoukian, A. and K. El Emam (2014). De-identification Protocols: Essential for Protecting Privacy. Privacy-By-Design Papers. Toronto, Canada, Information and Privacy Commisioner of Ontario.
Cavoukian, A. and D. Kruger (2014). Freedom and Control: Engineering a New Paradigm for the Digital World. Privacy-by-Design Papers, Information and Privacy Commissioner of Ontario, Canada.
Cavoukian, A., D. Stewart and B. Dewitt (2014). Using Privacy by Design to Achieve Big Data Innovation Without Compromising Privacy: Executive Summary, PbD & Deloitte.
Cavoukian, A. and A. Stoianov (2014). Privacy by Design Solutions for Biometric One-to-Many Identification Systems. Privacy-By-Design Papers. Toronto, Canada, Information and Privacy Commisioner of Ontario, Canada.
Cavusoglu, H., H. Cavusoglu, J.-Y. Son and I. Benbasat (2015). "Institutional pressures in security management: Direct and indirect influences on organizational investment in information security control resources." Information & Management **52**(4): 385-400.
Cavusoglu, H., B. Mishra and S. Raghunathan (2004). "A model for evaluating IT security investments." Communications of the ACM **47**(7): 87-92.
Ciriani, S. (2015). "The Economic Impact of the European Reform of Data Protection." Communications & Strategies(97): 41-58,153.
CISCO (2014). Cisco 2014 Annual Security Report**:** 1-80.
CL&SR (2013). "Data protection in Europe – Academics are taking a position." Computer Law & Security Review **29**(2): 180-184.
Clarke, R. (1988). "Information technology and dataveillance." Communications of the ACM **31**(5): 512.
Clarke, R. (1994). "The digital persona and its application to data surveillance." The information society **10**(2): 77-92.
Clarke, R. (1996) "Privacy and Dataveillance, and Organisational Strategy."
Clarke, R. (2006) "Introduction to Dataveillance and Information Privacy, and Definitions of Terms."
Clifton, C., M. Kantarcioglu and J. Vaidya (2002). Defining privacy for data mining, Citeseer.
Cockcroft, S. (2003). "Gaps between policy and practice in the protection of data privacy." JITTA **4**: 1.
CoE (1981). Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data. C. o. Europe.
Cohen, W. M. and S. Klepper (1996). "Firm Size and the Nature of Innovation within Industries: The Case of Process and Product R&D." The Review of Economics and Statistics **78**(2): 232-243.
Colonna, L. (2014). "Article 4 of the EU Data Protection Directive and the irrelevance of the EU–US Safe Harbor Program?" International Data Privacy Law **4**(3): 203-221.
Coltman, T., T. M. Devinney, D. F. Midgley and S. Venaik (2008). "Formative versus reflective measurement models: Two applications of formative measurement." Journal of Business Research **61**(12): 1250-1262.
Coll, S. (2014). "Power, knowledge, and the subjects of privacy: understanding privacy as the ally of surveillance." Information, Communication & Society **17**(10): 1250-1263.
Connolly, C. (2008). "The US Safe Harbor - Fact or Fiction?", from http://www.galexia.com/public/research/assets/safe_harbor_fact_or_fiction_2008/safe_harbor_fact_or_fiction.pdf.
Constantine, C. (2014). "Big data: an information security context." Network Security **2014**(1): 18-19.

Costello, A. B. and J. W. Osborne (2005). "Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis." Practical Assessment, Research & Evaluation **10**(7): 1-9.
Cramer, D. and D. Howitt (2004). Bartlett's test of sphericity. The SAGE Dictionary of Statistics. D. Cramer and D. Howitt. London, England, SAGE Publications, Ltd**:** 12.
Cronbach, L. J. (1951). "Coefficient Alpha and the Internal Structure of Tests." Psyckometrika **16**: 297-334.
Cronbach, L. J. and P. E. Meehl (1955). "Construct validity in psychological tests." Psychological Bulletin **52**(4): 281-302.
Crossler, R. E., A. C. Johnston, P. B. Lowry, Q. Hu, M. Warkentin and R. Baskerville (2013). "Future directions for behavioral information security research." Computers & Security **32**(0): 90-101.
Cudeck, R. and M. W. Browne (1983). "Cross-Validation Of Covariance Structures." Multivariate Behavioral Research **18**(2): 147-167.
Culnan, M. J. (1993). ""How Did They Get My Name?": An Exploratory Investigation of Consumer Attitudes Toward Secondary Information Use." MIS Quarterly **17**(3): 341-363.
Culnan, M. J. (1999). Georgetown Internet Privacy Policy Survey: Report to the Federal Trade Commission.
Culnan, M. J. (1999). Privacy and the Top 100 Websites: Report to the Federal Trade Commission, prepared for the Online Privacy Alliance.
Culnan, M. J. (2000). "Protecting Privacy Online: Is Self-Regulation Working?" Journal of Public Policy & Marketing **19**(1): 20-26.
Culnan, M. J. and P. K. Armstrong (1999). "Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation." Organization Science **10**(1): 104-115.
Culnan, M. J. and R. J. Bies (2003). "Consumer Privacy: Balancing Economic and Justice Considerations." Journal of Social Issues **59**(2): 323-342.
Culnan, M. J. and H. J. Smith (1995). Lotus Marketplace: Households–Managing Information Privacy Concerns. Computer Ethics and Social Values. D. G. Johnson and H. Nissenbaum, Prentice Hall.
Culnan, M. J. and C. C. Williams (2009). "How Ethics Can Enhance Organizational Privacy: Lessons From The ChoicePoint and TJX Breaches." MIS Quarterly **33**(4): 673-687.
Chai, S., M. Kim and H. R. Rao (2011). "Firms' information security investment decisions: Stock market evidence of investors' behavior." Decision Support Systems **50**(4): 651-661.
Chakravartty, A. (forthcoming). Scientific Realism. The Stanford Encyclopedia of Philosophy. E. N. Zalta.
Chan, Y., M. Culnan, K. Greenaway, G. Laden, T. Levin and H. J. Smith (2005). "Information Privacy: Management, Marketplace, And Legal Challenges." Communications of the Association for Information Systems(16): 270-298.
Chin, W. W. (1998). "Issues and opinion on structural equation modeling." MIS Quarterly **22**(1): VII-XVI.
Choi, B. C. K. and A. W. P. Pak (2005). "A catalog of biases in questionnaires." Preventing Chronic Disease **2**(1): 1-13.
ChoicePoint. (2009, 10/19/2009). "ChoicePoint, Inc. Agrees to Supplemental Provisions with Federal Trade Commission." from http://www.lexisnexis.com/risk/newsevents/press-release.aspx?Id=1258571228500519.
Choo, K.-K. R. (2011). "The cyber threat landscape: Challenges and future research directions." Computers & Security **30**(8): 719-731.
Christensen, L. and F. Etro (2013). "Big Data, the Cloud and the EU Regulation on Data Protection." Intereconomics **48**(5): 276-280.
Christopher Westland, J. (2010). "Lower bounds on sample size in structural equation modeling." Electronic Commerce Research and Applications **9**(6): 476-487.
Da Veiga, A. and J. Eloff (2010). "A framework and assessment instrument for Information Security Culture." Computers & Security **29**: 196-207.
Da Veiga, A. and N. Martins (2015). "Improving the information security culture through monitoring and implementation actions illustrated through a case study." Computers & Security **49**: 162-176.

Da Veiga, A. and N. Martins (2015). "Information security culture and information protection culture: A validated assessment instrument." Computer Law & Security Review **31**(2): 243-256.
Dale, K. L. C. R. M. (2015). "RIM's Role in Harnessing the Power of Big Data." Information Management **49**(4): 29-32,47.
Dateas. (2010). "Cómo Investigar Personas " Retrieved 2 May 2011, from http://www.dateas.com/es/como_investigar_personas.
Datta, A., M. C. Tschantz and A. Datta (2015). Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination. Proceedings of Privacy Enhancing Technologies Symposium, July 2015.
Davcik, N. S. (2014). "The use and misuse of structural equation modeling in management research." Journal of Advances in Management Research **11**(1): 47-81.
Davenport, T. H. (2006). "Competing On Analytics." Harvard Business Review **84**(1): 98-107.
Davenport, T. H. (2007). "BI and the Business Experiment." Business Intelligence Review: 11-23.
Davenport, T. H. (2009). "How to Design Smart Business Experiments." Harvard Business Review **87**(2): 68-76.
Davenport, T. H. (2014). Big Data @ Work: Dispelling the Myths, Uncovering the Opportunities. Boston: Massachusetts, Harvard Business Review Press.
Davenport, T. H. (2014). Big Data at Work: Dispelling the Myths, Uncovering the Opportunities, Harvard Business School Publishing Corp.
Davenport, T. H., P. Barth and R. Bean (2012). "How 'Big Data' Is Different. (cover story)." MIT Sloan Management Review **54**(1): 43-46.
Davenport, T. H. and J. Dyché (2013). Big Data in Big Companies, International Institute For Analytics**:** 1-31.
Davenport, T. H. and J. Dyché. (2013, May). "Big Data in Big Companies." from http://www.sas.com/reg/gen/corp/2266746.
Davenport, T. H., J. Harris and J. Shapiro (2010). "Competing on talent analytics." Harvard Business Review **88**(10): 52-58.
Davenport, T. H. and J. G. Harris (2005). "Automated Decision Making Comes of Age." MIT Sloan Management Review **46**(4): 83-89.
Davenport, T. H. and J. G. Harris (2007). Competing on Analytics: The New Science of Winning. Boston, MA, Harvard Business School Press.
Davenport, T. H. and J. G. Harris (2007). "Competing with Multichannel Marketing Analytics." Advertising Age **78**(14): 16-17.
Davenport, T. H. and J. G. Harris (2009). "What People Want (and How to Predict It)." MIT Sloan Management Review **50**(2): 23-31.
Davenport, T. H. and J. G. Harris (2010). "Leading the way towards better business insights." Strategic HR Review **9**(4): 28-33.
Davenport, T. H., J. G. Harris, D. W. De Long and A. L. Jacobson (2001). "Data to Knowledge to Results: Building an Analytic Capability." California Management Review **43**(2): 117-138.
Davenport, T. H., J. G. Harris, D. W. De Long and A. L. Jacobson (2001). "Data to Knowledge to Results: Bulding an Analytic Capability." California Management Review **43**(2): 117-138.
Davenport, T. H., J. G. Harris, G. L. Jones, K. N. Lemon, D. Norton and M. B. McCallister (2007). "The Dark Side of Customer Analytics." Harvard Business Review **85**(5): 37-48.
Davenport, T. H., J. G. Harris and R. Morison (2010). Analytics at Work: Smarter Decisions, Better Results. Boston (MA), Harvard Business Press.
Davenport, T. H., L. D. Mule and J. Lucker (2011). "Know What Your Customers Want Before They Do." Harvard Business Review **89**(12): 84-92.
Davenport, T. H. and D. J. Patil (2012). "Data Scientist: The Sexiest Job Of the 21st Century." Harvard Business Review **90**(10): 70-76.
Davey, N. (2010). "Data industry reaches landmark - but insight remains elusive." News Retrieved March 11, 2011, 2011, from http://www.knowledgeboard.com/item/3070/23/5/3.
Davison, R. M. and M. G. Martinsons (2015). "Context is king! Considering particularism in research design and reporting." Journal of Information Technology.
Dean, T. J. and R. L. Brown (1995). "Pollution regulation as a barrier to new firm entry: Initial." Academy of Management Journal **38**(1): 288.

Degli-Esposti, S. (2012). "Aumentar la seguridad de la información mediante el respecto a la privacidad: algunos ejemplos " Novática **218**: 65-69.
Degli Esposti, S. (2014). "When big data meets dataveillance: The hidden side of analytics." Surveillance & Society **12**(2): 209-225.
Deloitte (2007). Enterprise@Risk: Insights into the emerging privacy and data protection function - 2007 Privacy & Data Protection Survey. Audit & Enterprise Risk Services, Deloitte & Touche LLP and Ponemon Institute LLC. **7424:** 40.
DG-Justice. (2012). "Commission proposes a comprehensive reform of the data protection rules." from http://ec.europa.eu/justice/newsroom/data-protection/news/120125_en.htm.
Dhillon, G., R. Syed and C. Pedron (2016). "Interpreting information security culture: An organizational transformation case study." Computers & Security **56**: 63-69.
DHS (2008). Privacy Policy Guidance Memorandum. H. T. III, Privacy Office of the U.S. Department of Homeland Security.
Diamantopoulos, A. (2005). "The C-OAR-SE procedure for scale development in marketing: a comment." International Journal of Research in Marketing **22**(1): 1-9.
Diamantopoulos, A., P. Riefler and K. P. Roth (2008). "Advancing formative measurement models." Journal of Business Research **61**(12): 1203-1218.
Diamantopoulos, A. and H. M. Winklhofer (2001). "Index Construction with Formative Indicators: An Alternative to Scale Development." Journal of Marketing Research (JMR) **38**(2): 269-277.
Dibb, S., K. Ball, A. Canhoto, E. M. Daniel, M. Meadows and K. Spiller (2014). "Taking responsibility for border security: Commercial interests in the face of e-borders." Tourism Management **42**: 50-61.
Dickson, J. and G. Albaum (1977). "A Method for Developing Tailormade Semantic Differentials for Specific Marketing Content Areas." Journal of Marketing Research (JMR) **14**(1): 87-91.
Dienlin, T. and S. Trepte (2015). "Is the privacy paradox a relic of the past? An in-depth analysis of privacy attitudes and privacy behaviors." European Journal of Social Psychology **45**(3): 285-297.
Dinev, T. and P. Hart (2004). Internet privacy concerns and their antecedents - measurement validity and a regression model. Behaviour & Information Technology, Taylor & Francis Ltd. **23:** 413-422.
Dix, A., G. Thüsing, J. Traut, L. Christensen, F. Etro, S. A. Aaronson and R. Maxim (2013). "EU data protection reform: Opportunities and concerns." Intereconomics **48**(5): 268-285.
Dolch, N. A. (1980). "Attitude Measurement by Semantic Differential on a Bipolar Scale." Journal of Psychology **105**(2): 151.
Dolnicar, S. and Y. Jordaan (2007). "A MARKET-ORIENTED APPROACH TO RESPONSIBLY MANAGING INFORMATION PRIVACY CONCERNS IN DIRECT MARKETING." Journal of Advertising **36**(2): 123-149.
Doornik, J. A. and H. Hansen (2008). "An Omnibus Test for Univariate and Multivariate Normality." Oxford Bulletin of Economics & Statistics **70**: 927-939.
Dow, K. E., J. Wong, C. Jackson and R. A. Leitch (2008). "A comparison of structural equation modeling approaches: the case of user acceptance of information systems." The Journal of Computer Information Systems **48**(4): 106-114.
Dowling, G. R. (2004). "Journalists' evaluation of corporate reputations." Corporate Reputation Review **7**(2): 196–205.
Du, D., A. Li and L. Zhang (2014). "Survey on the Applications of Big Data in Chinese Real Estate Enterprise." Procedia Computer Science **30**: 24-33.
Dutta, A. and K. McCrohan (2002). "Management's Role in Information Security in a Cyber Economy." California Management Review **45**(1): 67-87.
Earp, J. B., A. I. Antón, L. Aiman-Smith and W. H. Stufflebeam (2005). "Examining Internet Privacy Policies Within the Context of User Privacy Values." IEEE Transactions on Engineering Management **52**(2): 227-237.
Easton, G. (2010). "Critical realism in case study research." Industrial Marketing Management **39**(1): 118-128.
EC-R-45 (2001). "REGULATION (EC) No 45/2001 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 18 December 2000 on the protection of individuals with regard to

the processing of personal data by the Community institutions and bodies and on the free movement of such data." Official Journal of the European Communities **8**: 1-22.
EC (1995). Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, Official Journal L 281**:** 31-50.
EC (1996). Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases Official Journal. **77:** 20-28.
EC (1999). Directive 1999/5/EC of the European Parliament and of the Council of 9 March 1999 on radio equipment and telecommunications terminal equipment and the mutual recognition of their conformity, Official Journal**:** 10-28.
EC (2014). Progress on EU data protection reform now irreversible following European Parliament vote. MEMO, European Commision.
EC/21 (2002). "DIRECTIVE 2002/21/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 7 March 2002 on a common regulatory framework for electronic communications networks and services (Framework Directive)." Official Journal of the European Communities **108**: 33-50.
EC/22 (2002). "DIRECTIVE 2002/22/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 7 March 2002 on universal service and users' rights relating to electronic communications networks and services (Universal Service Directive)." Official Journal of the European Communities **108**: 51-77.
EC/24 (2006). "DIRECTIVE 2006/24/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC." Official Journal of the European Communities **105**(EN): 54-63.
EC/31 (2000). "DIRECTIVE 2000/31/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce)." Official Journal of the European Communities **78**: 1-16.
EC/46 (1995). Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, Official Journal L 281**:** 31-50.
EC/58 (2002). "DIRECTIVE 2002/58/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications)." Official Journal of the European Communities **201**(EN): 37-47.
EC/66 (1997). "DIRECTIVE 97/66/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 15 December 1997 concerning the processing of personal data and the protection of privacy in the telecommunications sector." Official Journal of the European Communities **24**: 1-8.
EC/136 (2009). "DIRECTIVE 2009/136/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 25 November 2009 amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services, Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector and Regulation (EC) No 2006/2004 on cooperation between national authorities responsible for the enforcement of consumer protection laws (Text with EEA relevance)." Official Journal of the European Communities **337**(EN): 11-36.
ECDP (2010). Summary of replies to the public consultation about the future legal framework for protectiong personal data. Brussels, European Commission; Directorate-General Justice; Directorate C: Fundamental rights and Union citizenship; Unit C.3: Data protection.
Edwards, J. R. and R. P. Bagozzi (2000). "On the nature and direction of relationships between constructs and measures." Psychological Methods **5**(2): 155-174.
Eloff, M. M. and S. H. von Solms (2000). "Information Security Management: An Approach to Combine Process Certification And Product Evaluation." Computers & Security **19**(8): 698-709.
Engau, C. and V. H. Hoffmann (2011). "Corporate response strategies to regulatory uncertainty: evidence from uncertainty about post-Kyoto regulation." Policy Sciences **44**(1): 53-80.

EU-FRA and CoU (2014). Handbook on European data protection law, Council of Europe and the European Union Agency for Fundamental Rights**:** 1-199.
EU (2000). "Charter Of Fundamental Rights Of The European Union." Official Journal of the European Communities **364**(EN).
EU (2007). "Treaty of Lisbon amending the Treaty on European Union and the Treaty establishing the European Community, signed at Lisbon, 13 December 2007." Official Journal of the European Union **50**(C 306): 1-271.
Fabrigar, L. R., R. D. Porter and M. E. Norris (2010). "Some things you should know about structural equation modeling but never thought to ask." Journal of Consumer Psychology **20**(2): 221-225.
Fabrigar, L. R., D. T. Wegener, R. C. MacCallum and E. J. Strahan (1999). "Evaluating the use of exploratory factor analysis in psychological research." Psychological Methods **4**(3): 272-299.
Featherman, M. S., A. D. Miyazaki and D. E. Sprott (2010). "Reducing online privacy risk to facilitate e-service adoption: the influence of perceived ease of use and corporate credibility." Journal of Services Marketing **24**(3): 219-229.
Ferguson, R. B. (2012). "Risky Business: How Data Analytics and Behavioral Science Can Help." MIT Sloan Management Review **54**(1): 1-5.
Ferguson, R. B. (2013). "How eBay Uses Data and Analytics to Get Closer to Its (Massive) Customer Base." MIT Sloan Management Review **55**(1): 1-3.
Field, A. (2013). "Moderation and Mediation." from https://www.youtube.com/watch?v=RqkGMqDU20Q.
Fink, A. G. (2002). How to Design Survey Studies, Sage Publications, Inc.
Finn, A. and U. Kayande (2005). "How fine is C-OAR-SE? A generalizability theory perspective on Rossiter's procedure." International Journal of Research in Marketing **22**(1): 11-21.
Flaherty, D. H. (1989). Protecting privacy in surveillance societies. Chapel Hill, NC, USA, University of North Carolina Press.
Flaherty, D. H. (1989). Protecting privacy in surveillance societies: The Federal Republic of Germany, Sweden, France, Canada, and the United States. Chapel Hill, NC, The University of North Carolina Press.
Flavelle, D. (2010) "What the data crunchers know about you."
Fleiss, J. L. and J. Cohen (1973). "The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability." Educational and Psychological Measurement **33**(3): 613-619.
Florencio, D. and C. Herley. (2011, June 2011). "Sex, Lies and Cyber-crime Surveys." Retrieved 13/04/2015, from http://research.microsoft.com/pubs/149886/SexLiesandCybercrimeSurveys.pdf.
Fogarty, D. and P. C. Bell (2014). "Should You Outsource Analytics?" MIT Sloan Management Review **55**(2): 41-45.
Fortin, S. and B. M. Knoppers (2009). "Secondary Uses of Personal Data for Population Research." Genomics, Society & Policy **5**(1): 80-99.
Fox, J. (1980). "Effect Analysis in Structural Equation Models: Extensions and Simplified Methods of Computation." Sociological Methods & Research **9**(1): 3-28.
Fox, S. (2009). "Applying critical realism to information and communication technologies: a case study." Construction Management & Economics **27**(5): 465-472.
Foxman, E. R. and P. Kilcoyne (1993). "Information Technology, Marketing Practice, and Consumer Privacy: Ethical Issues." Journal of Public Policy & Marketing **12**(1): 106-119.
Franke, U. and J. Brynielsson (2014). "Cyber situational awareness – A systematic review of the literature." Computers & Security **46**(0): 18-31.
Frasher, M. (2013). "Adequacy versus equivalency: Financial data protection and the U.S.–EU divide." Business Horizons **56**(6): 787-795.
FTC (2000). Fair Information Practices in the Electronic Marketplace, US Federal Trade Commission.
FTC (2012). Fair Credit Reporting Act: Complete Text, Federal Trade Commission.
FTC (2012). Protecting Consumer Privacy in an Era of Rapid Change: Recommendations For Businesses and Policymakers, FEDERAL TRADE COMMISSION.
Furnell, S. M. and M. J. Warren (1999). "Computer hacking and cyber terrorism: the real threats in the new millennium?" Computers & Security **18**(1): 28-34.

Gandy, O. (1993). The panoptic sort: A political economy of personal information, Westview Press CO.
Gandy, O. H. (2003). "Public Opinion Surveys and the Formation of Privacy Policy." Journal of Social Issues **59**(2): 283-299.
Gandy, O. H. (2010). "Engaging rational discrimination: exploring reasons for placing regulatory constraints on decision support systems." Ethics and information technology **12**(1): 29-42.
Gandy, O. H. (2012). 2.1.a. Statistical Surveillance: Remote Sensing in the Digital Age. Handbook of Surveillance Studies. K. Ball, K. Haggerty and D. Lyon. New York, Routledge**:** 125-132.
Gantz, J. and D. Reinsel (2010). The Digital Universe Decade – Are You Ready? Framingham, MA 01701 USA, EMC Corporation.
GAO, U. S. A. (2007). Personal Information: Data Breaches Are Frequent, but Evidence of Resulting Identity Theft Is Limited; However, the Full Extent Is Unknown. Report to Congressional Requesters, U.S. Government Accountability Office**:** 1-37.
Gartner. (2013). "IT Glossary." from http://www.gartner.com/it-glossary/big-data/.
Gilliom, J. (2001). Overseers of the Poor: Surveillance, Resistance, and the Limits of Privacy, The University of Chicago Press.
Gilliom, J. (2011). "A Response to Bennett's' In Defence of Privacy'." Surveillance & Society **8**(4): 500-504.
Girotra, K. and S. Netessine (2014). "Four paths to business model innovation." Harvard Business Review **92**(7/8): 96-103.
Goes, P. B. (2014). "Big Data and IS Research." MIS Quarterly **38**(3): iii-viii.
Google. (2015). "Google Trends." from http://www.google.com/trends/explore#q=%2Fm%2F016jq3%2C%20%2Fm%2F0bs2j8q%2C%20%2Fm%2F0blvg%2C%20%2Fm%2F02gcn9&cmpt=q.
Google. (2015). "Google's mission is to organize the world's information and make it universally accessible and useful.", from http://www.google.com/about/company/.
Gordon, L. A. and M. P. Loeb (2002). "The economics of information security investment." ACM Trans. Inf. Syst. Secur. **5**(4): 438-457.
Gordon, L. A. and M. P. Loeb (2002). "Return on information security investments: Myths vs. realities." Strategic Finance **84**(5): 26-31.
Gordon, L. A. and M. P. Loeb (2006). "Economic aspects of information security: An emerging field of research." Information Systems Frontiers **8**(5): 335.
Goucher, W. (2011). "Do SMEs have the right attitude to security?" Computer Fraud & Security **2011**(7): 18-20.
Grapentine, T. (2000). "Path analysis vs. structural equation modeling." Marketing Research **12**(3): 12-19.
Greenaway, K. E. and Y. E. Chan (2005). Theoretical Explanations for Firms' Information Privacy Behaviors. Journal of the Association for Information Systems, Association for Information Systems. **6:** 171-189.
Greenaway, K. E. and Y. E. Chan (2005). "Theoretical Explanations for Firms' Information Privacy Behaviors." Journal of the Association for Information Systems **6**: 171-189.
Greenaway, K. E. and Y. E. Chan (2013). "Designing a Customer Information Privacy Program Aligned with Organizational Priorities." MIS Quarterly Executive **12**(3): 137-150.
Grubmüller, V., K. Götsch and B. Krieger (2013). "Social media analytics for future oriented policy making." European Journal of Futures Research **1**(1): 1-9.
Grundvig, J. (2014) "Does Facebook Know Your Location? Not for Long With Location Sentry's App." Huffington Post - The Blog: Featuring fresh takes and real-time analysis from HuffPost's signature lineup of contributors.
Guba, E. G. (1990). The alternative paradigm dialog. The paradigm dialog. E. G. Guba**:** 17-30.
Guilford, J. P. (1941). "The phi coefficient and chi square as indices of item validity." Psychometrika **6**(1): 11-19.
Guilford, J. P. (1948). "Factor analysis in a test-development program." Psychological Review **55**(2): 79-94.
Gürses, S. and B. Berendt (2010). PETs in the Surveillance Society: A Critical Review of the Potentials and Limitations of the Privacy as Confidentiality Paradigm. Data Protection in a

Profiled World. S. Gutwirth, Y. Poullet and P. De Hert. Dordrecht Heidelberg London New York, Springer.
Gutwirth, S., Y. Poullet, P. De Hert and R. Leenes, Eds. (2011). Computers, Privacy and Data Protection: an Element of Choice. Dordrecht Heidelberg London New York, Springer
Hacking, I. (1982). "Experimentation and Scientific Realism." Philosophical Topics **13**(1): 71-87.
Hagel, J. and J. F. Rayport (1997). "The coming battle for customer information." McKinsey Quarterly: 64-77.
Haggerty, K. D. and R. V. Ericson (2000). "The surveillant assemblage." The British Journal of Sociology **51**(4): 605-622.
Haggerty, N. (2012). "On becoming an IT savvy CEO." Ivey Business Journal **76**(4): 1-4.
Hair, J. F., W. C. Black, B. J. Babin and R. E. Anderson (2009). Multivariate Data Analysis, Prentice Hall.
Hamid, I. R. A. and J. H. Abawajy (2014). "An approach for profiling phishing activities." Computers & Security **45**(0): 27-41.
Hardin, S. (2015). "Alessandro Acquisti addresses ASIS&T plenary session." Bulletin of the American Society for Information Science and Technology **41**(3): 33-35.
Hardt, M. and S. Nath (2012). Privacy-aware personalization for mobile advertising. Proceedings of the 2012 ACM conference on Computer and communications security. Raleigh, North Carolina, USA, ACM**:** 662-673.
Harris, J. G. and T. H. Davenport (2007). "Competing on Analytics: The New Science of Winning." Harvard Business School Press Books: 1.
Harris, J. G., R. Morison and T. H. Davenport (2010). "Analytics at Work: Smarter Decisions, Better Results." Harvard Business School Press Books: 1.
Hashem, I. A. T., I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani and S. Ullah Khan (2015). "The rise of "big data" on cloud computing: Review and open research issues." Information Systems **47**: 98-115.
Hayes, A. F. (2009). "Beyond Baron and Kenny: Statistical Mediation Analysis in the New Millennium." Communication Monographs **76**(4): 408-420.
Hayes, A. F. (2013). Introduction to Mediation, Moderation, and Conditional Process Analysis : A Regression-Based Approach. Methodology in the Social Sciences. G. Press, The Guilford Press.
Helm, S. (2005). "Designing a formative measure for corporate reputation." Corporate Reputation Review **8**(2): 95–109.
Heng, X., T. Dinev, J. Smith and P. Hart (2011). "Information Privacy Concerns: Linking Individual Perceptions with Institutional Privacy Assurances." Journal of the Association for Information Systems **12**(12): 798-824.
Henze, N. and B. Zirkler (1990). "A class of invariant consistent tests for multivariate normality." Communications in Statistics - Theory and Methods **19**(10): 3595-3617.
HEW (1973). Records, Computers and the Rights of Citizens Report of the Secretary's Advisory Committee on Automated Personal Data Systems. E. U.S. Department of Health, and Welfare. Washington, DC.
HEW (1973). Records, Computers and the Rights of Citizens: Report of the Secretary's Advisory Committee on Automated Personal Data Systems, United States Department of Health, Education and Welfare.
Hilbe, J. M. (2014). Chapter 2. Review of Generalized Linear Models and Generalized Estimating Equations. Quasi-Least Squares Regression. J. M. Hilbe, Chapman and Hall/CRC**:** 17-40.
Hitt, L. M. and E. Brynjolfsson (1996). "Productivity, Business Profitability, and Consumer Surplus: Three Different Measures of Information Technology Value." MIS Quarterly **20**(2): 121-142.
Hitt, L. M. and E. Brynjolfsson (1997). "Information Technology and Internal Firm Organization: An Exploratory Analysis." Journal of Management Information Systems **14**(2): 81-101.
Hoffer, J. A. and D. W. Straub (1989). "The 9 to 5 Underground: Are You Policing Computer Crimes?" Sloan Management Review **30**(4): 35.
Hofstede, G. (1980). Culture's Consequences: International Differences in Work-Related Values. Beverly Hills, CA, Sage.
Hofstede, G. (1991). Cultures and Organizations. Berkshire, UK, McGraw-Hill.

Hollander, M., D. A. Wolfe and E. Chicken (2014). Nonparametric statistical methods, John Wiley & Sons.
Homburg, C., W. D. Hoyer and M. Fassnacht (2002). "Service Orientation of a Retailer's Business Strategy: Dimensions, Antecedents, and Performance Outcomes." Journal of Marketing **66**(4): 86-101.
Homburg, C., J. P. Workman, Jr. and H. Krohmer (1999). "Marketing's Influence within the Firm." Journal of Marketing **63**(2): 1-17.
Hoofnagle, C. J. (2004). "Big Brother's Little Helpers: How ChoicePoint and Other Commercial Data Brokers Collect and Package Your Data for Law Enforcement." North Carolina Journal of International Law and Commercial Regulation **29**(4): 595-638.
Hoofnagle, C. J. (2007). Security Breach Notification Laws: Views from Chief Security Officers, A Study Conducted for the Samuelson Law, Technology & Public Policy Clinic, University of California-Berkeley School of Law.
Hoofnagle, C. J. (2010). Country Studies: B.1 - United States Of America. Comparative Study On Different Approaches To New Privacy Challenges, In Particular In The Light Of Technological Developments. D. Korff, LRDP KANTOR Ltd & The Centre for Public Reform.
Hoofnagle, C. J., A. Soltani, N. Good, D. J. Wambach and M. D. Ayenson (2012). "Behavioral Advertising: The Offer You Cannot Refuse." Harvard Law & Policy Review **6**(2): 273-296.
Hoy, M. G. and J. Phelps (2003). "Consumer Privacy and Security Protection on Church Web Sites: Reasons for Concern." Journal of Public Policy & Marketing **22**(1): 58-70.
Hubert, L. and H. Wainer (2012). The Basic Sampling Model and Associated Topics. A Statistical Guide for the Ethically Perplexed. L. Hubert and H. Wainer, Chapman and Hall/CRC**:** 175-225.
Hughes, S. S. (2012). "US Domestic Surveillance after 9/11: An Analysis of the Chilling Effect on First Amendment Rights in Cases Filed against the Terrorist Surveillance Program." Canadian journal of law and society **27**(3): 399-425.
Hui-Chih, W. and D. Her-Sen (2010). "Nine issues for Internet-based survey research in service industries." Service Industries Journal **30**(14): 2387-2399.
Hui, K.-L., H. H. Teo and S.-Y. Tom Lee (2007). The Value of Privacy Assurance: An Exploratory Field Experiment. MIS Quarterly, MIS Quarterly & The Society for Information Management. **31:** 19-33.
Hulin, C. and R. Cudeck (2001). "Cronbach's Alpha on Two-Item Scales." Journal of Consumer Psychology **10**(1/2): 55.
Iacobucci, D. (2010). "Structural equations modeling: Fit Indices, sample size, and advanced topics." Journal of Consumer Psychology **20**(1): 90-98.
IAPP (2010). A Call For Agility: The Next-Generation Privacy Professional. York, ME, International Association of Privacy Professionals.
IBM (2013). What will we make of this moment? 2013 IBM Annual Report.
ICO. (2013). "Your Thoughts: Join the Open University's 'Big Data Protection' Study"." from http://ico.msgfocus.com/q/1bkJRNvP4UxkK5O3bL5/wv.
ICO. (2014). "Last chance to join the Open University's 'Big Data Protection' Study." from http://ico.msgfocus.com/q/1AFxOgho8z/wv.
ICO (2014). Review of the impact of ICO Civil Monetary Penalties.
Il-Horn, H., H. U. I. Kai-Lung, L. E. E. Sang-Yong Tom and I. P. L. Png (2007). Overcoming Online Information Privacy Concerns: An Information-Processing Theory Approach. Journal of Management Information Systems, M.E. Sharpe Inc. **24:** 13-42.
Inagaki, K. (2014). Sony PlayStation online store hacked. The Financial Times. FT.com, The Financial Times Limited.
Inkster, N. (2015). The Chinese Intelligence Agencies: Evolution and Empowerment in Cyberspace. China and Cybersecurity: Espionage, Strategy, and Politics in the Digital Domain. J. R. Lindsay, T. M. Cheung and D. S. Reveron. New York, NY, Oxford University Press, Apr 7, 2015**:** 29-50.
ISFE (2010). European Commission Communication on A Comprehensive Approach to Personal Data Protection In the European Union - ISFE Statement. **Interactive Software Federation of Europe (ISFE)**.
Iyer, B. and T. H. Davenport (2008). "Reverse Engineering Google's Innovation Machine. (cover story)." Harvard Business Review **86**(4): 58-68.

Jeffrey T. Steedle, R. J. S. (2010). “Coefficient Alpha and the Internal Structure of Tests“. Encyclopedia of Research Design. SAGE Publications, Inc. Encyclopedia of Research Design. N. J. Salkind. Thousand Oaks, CA, SAGE Publications, Inc.**:** 164-165.
Jehn-Yih, W. and C. Pi-Heng (2008). "Retaining Passenger Loyalty through Data Mining: A Case Study of Taiwanese Airlines." Transportation Journal **47**(1): 17-29.
Jensen, C. and C. Potts (2004). Privacy policies as decision-making tools: an evaluation of online privacy notices. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Vienna, Austria, ACM**:** 471-478.
Johansson, J. K. and G. S. Yip (1994). "Exploiting globalization potential: U.S. and Japanese strategies." Strategic Management Journal **15**(8): 579-601.
John, L. K., A. Acquisti and G. Loewenstein (2011). "Strangers on a Plane: Context-Dependent Willingness to Divulge Sensitive Information." The Journal of Consumer Research **37**(5): 858-873.
Johnson, A. M. and B. P. Shipps (2013). "Acquiring Subject Participation for Information Security Survey Research: A Content and Correspondence Analysis Approach." Journal of Information Privacy & Security **9**(4): 3-30.
Jones, M. G. (1991). "Privacy: A Significant Marketing Issue for the 1990s." Journal of Public Policy & Marketing **10**(1): 133-148.
Kaiser, H. (1958). "The varimax criterion for analytic rotation in factor analysis." Psychometrika **23**(3): 187-200.
Kaiser, H. (1974). "An index of factorial simplicity." Psychometrika **39**(1): 31-36.
Kambatla, K., G. Kollias, V. Kumar and A. Grama (2014). "Trends in big data analytics." Journal of Parallel and Distributed Computing **74**(7): 2561-2573.
Kane, M. J. and D. A. Ricks (1988). "Is Transnational Data Flow Regulation a Problem?" Journal of International Business Studies **19**(3): 477-482.
Keen, P. and R. Williams (2013). "Value architectures for digital business: Beyond the business model." MIS Quarterly **37**(2): 643-647.
Kehr, F., T. Kowatsch, D. Wentzel and E. Fleisch (2015). "Blissfully ignorant: the effects of general privacy concerns, general institutional trust, and affect in the privacy calculus." Information Systems Journal: n/a-n/a.
Kelley, P. G., L. F. Cranor and N. Sadeh (2013). Privacy as part of the app decision-making process. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Paris, France, ACM**:** 3393-3402.
Khansa, L. and D. Liginlal (2007). The Influence of regulations on innovation in information security.
Kieke, R. L. (2014). "Recent Privacy Breach Settlements Illustrate the Importance of Proactively Pursuing Compliance." Journal of Health Care Compliance **16**(4): 41-61.
Kiron, D., R. B. Ferguson and P. Kirk Prentice (2013). "From Value to Vision: Reimagining the Possible with Data Analytics." MIT Sloan Management Review **54**(3): 1-n/a.
Kiron, D., P. K. Prentice and R. B. Ferguson (2014). "The Analytics Mandate." MIT Sloan Management Review **55**(4): 1-25.
Knapp, E. D. and J. T. Langill (2015). Industrial Network Security: Securing Critical Infrastructure Networks for Smart Grid, SCADA, and Other Industrial Control Systems. Waltham (MA), Syngress-Elsevier.
Kobrin, S. J. (2004). "Safe harbours are hard to find: the trans-Atlantic data privacy dispute, territorial jurisdiction and global governance." Review of International Studies **30**(01): 111-131.
Kong, L. (2010). "Data Protection and Transborder Data Flow in the European and Global Context." European Journal of International Law **21**(2): 441-456.
Korff, D. (2010). Working Paper Nr. 2 - Data protection laws in the EU: The difficulties in meeting the challenges posed by global social and technical developments. Comparative Study On Different Approaches To New Privacy Challenges, In Particular In The Light Of Technological Developments, LRDP KANTOR Ltd & The Centre for Public Reform.
Kosinski, M., D. Stillwell and T. Graepel (2013). "Private traits and attributes are predictable from digital records of human behavior." Proceedings of the National Academy of Sciences **110**(15): 5802-5805.
Kotulic, A. G. and J. G. Clark (2004). "Why there aren’t more information security research studies." Information & Management **41**(5): 597-607.

Kruskal, W. H. and W. A. Wallis (1952). "Use of Ranks in One-Criterion Variance Analysis." Journal of the American Statistical Association **47**(260): 583-621.
Kruskal, W. H. and W. A. Wallis (1953). "Errata: Use of Ranks in One-Criterion Variance Analysis." Journal of the American Statistical Association **48**(264): 907-911.
Kshetri, N. (2006). "The simple economics of cybercrimes." Security & Privacy, IEEE **4**(1): 33-39.
Kshetri, N. (2013). "Cybercrimes in the Former Soviet Union and Central and Eastern Europe: current status and key drivers." Crime, Law and Social Change **60**(1): 39-65.
Kshetri, N. (2014). "Big data′s impact on privacy, security and consumer welfare." Telecommunications Policy **38**(11): 1134-1145.
Kuo, F.-Y., C. Lin and M.-H. Hsu (2007). "Assessing Gender Differences in Computer Professionals' Self-Regulatory Efficacy Concerning Information Privacy Practices." Journal of Business Ethics **73**(2): 145-160.
Kurpjuhn, T. (2015). "The SME security challenge." Computer Fraud & Security **2015**(3): 5-7.
Lace, S. (2005). The Glass Consumer: Life in a surveillance society. Bristol, The Policy Press.
Laney, D. (2001). "3D Data Management: Controlling Data Volume, Velocity, and Variety." META Group.
Lanier, C. D. and A. Saini (2008). "Understanding consumer privacy: a review and future directions." Academy of Marketing Science Review **12**(2): 1-48.
LaValle, S., E. Lesser, R. Shockley, M. S. Hopkins and N. Kruschwitz (2011). "Big Data, Analytics and the Path From Insights to Value." MIT Sloan Management Review **52**(2): 21-32.
Lecuyer, M., R. Spahn, Y. Spiliopolous, A. Chaintreau, R. Geambasu and D. Hsu (2015). Sunlight: Fine-grained Targeting Detection at Scale with Statistical Confidence. Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. Denver, Colorado, USA, ACM**:** 554-566.
Lee, D.-J., J.-H. Ahn and Y. Bang (2011). "Managing Consumer Privacy Concerns in Personalization: A Strategic Analysis of Privacy Protection." MIS Quarterly **35**(2): 423-A428.
Lee, Y. J., R. J. Kauffman and R. Sougstad (2011). "Profit-maximizing firm investments in customer information security." Decision Support Systems **51**(4): 904-920.
Leon, P. G., B. Ur, Y. Wang, M. Sleeper, R. Balebako, R. Shay, L. Bauer, M. Christodorescu and L. F. Cranor (2013). What matters to users?: factors that affect users' willingness to share information with online advertisers. Proceedings of the Ninth Symposium on Usable Privacy and Security. Newcastle, United Kingdom, ACM**:** 1-12.
Lester, T. (2001). "The Reinvention of Privacy: It used to be that business and technology were considered the enemies of privacy. Not anymore." The Atlantic Monthly **287**(3): 27-39.
Levi, M. and D. S. Wall (2004). "Technologies, Security, and Privacy in the Post-9/11 European Information Society." Journal of Law and Society **31**(2): 194–220.
Lewington, J., L. De Chernatony and A. Brown (1996). "Harnessing the Power of Database Marketing." Journal of Marketing Management **12**(4): 329-346.
Li, D. C. (2015). "Online security performance and information security disclousures." Journal of Computer Information Systems **55**(2): 20-28.
LIBE (2015). EU General Data Protection Regulation State of play and 10 main issues. Jan Philipp Albrecht, MdEP webpage, Lead EP Committee: Committee on Civil Liberties, Justice and Home Affairs (LIBE)**:** 1-3.
Lindsay, J. R., T. M. Cheung and D. S. Reveron, Eds. (2015). China and Cybersecurity: Espionage, Strategy, and Politics in the Digital Domain. New York, NY, Oxford University Press, Apr 7, 2015.
Linebach, J. A., B. P. Tesch and L. M. Kovacsiss (2014). Glossary. Nonparametric Statistics for Applied Research. J. A. Linebach, B. P. Tesch and L. M. Kovacsiss, Springer**:** 335-408.
Linebach, J. A., B. P. Tesch and L. M. Kovacsiss (2014). Nonparametric Statistics for Applied Research. New York, Springer.
Lipton, J. D. (2001). "Security Interests in Electronic Databases." International Journal of Law & Information Technology **9**(65).
Lohr, S. L. (2009). Multiple-Frame Surveys. Sample Surveys: Design, Methods and Applications, Vol. 29A. D. Pfeffermann and C. R. Rao, Elsevier**:** 71-88.
Loveman, G. (2003). "Diamonds in the Data Mine." Harvard Business Review **81**(5): 109-113.

Lyon, D. (1993). "An electronic panopticon? A sociological critique of surveillance theory." Sociological Review **41**(4): 653-678.
Lyon, D. (1994). The electronic eye: The rise of surveillance society. Minneapolis, University Of Minnesota Press.
Lyon, D. (2001). "Facing the future: Seeking ethics for everyday surveillance." Ethics and information technology **3**(3): 171-180.
Lyon, D. (2001). Surveillance society: Monitoring everyday life. Buckingham, Philadelphia, Open University Press.
Lyon, D. (2002). "Everyday Surveillance: Personal data and social classifications." Information, Communication & Society **5**(2): 242-257.
Lyon, D. (2007). "Surveillance, Security and Social Sorting." International criminal justice review **17**(3): 161.
Lyon, D. and E. Zureik (1996). "Surveillance, privacy and the new technology." Computers, surveillance, and privacy: 1–18.
MacCallum, R. C., K. F. Widaman, S. Zhang and S. Hong (1999). "Sample size in factor analysis." Psychological Methods **4**(1): 84-99.
MacKenzie, S. B., P. M. Podsakoff and C. B. Jarvis (2005). "The Problem of Measurement Model Misspecification in Behavioral and Organizational Research and Some Recommended Solutions." Journal of Applied Psychology **90**(4): 710-730.
MacKenzie, S. B., P. M. Podsakoff and N. P. Podsakoff (2011). "Construct measurement and validation procedures in MIS and Behavioral Research: Integrating new and existing techniques." MIS Quarterly **35**(2): 293-A295.
Madden, S. (2012). "From Databases to Big Data." IEEE Internet Computing **3**(16): 4-6.
Maillart, T. and D. Sornette (2010). "Heavy-tailed distribution of cyber-risks." European Physical Journal B -- Condensed Matter **75**(3): 357-364.
Malhotra, N., J. A. Krosnick and R. K. Thomas (2009). "Optimal Design of Branching Questions to Measure Bipolar Constructs." Public Opinion Quarterly **73**(2): 304-324.
Malhotra, N. K., S. K. Sung and J. Agarwal (2004). "Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model." Information Systems Research **15**(4): 336-355.
Mantelero, A. (2014). "The future of consumer data protection in the E.U. Re-thinking the "notice and consent" paradigm in the new era of predictive analytics." Computer Law & Security Review **30**(6): 643-660.
Manyika, J., M. Chui, J. Bughin, B. Brown, R. Dobbs, C. Roxburgh and A. Hung Byers (2011). Big data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute (MGI).
Mardia, K. V. (1970). "Measures of multivariate skewness and kurtosis with applications." Biometrika **57**(3): 519-530.
Margulis, S. T. (2003). "On the Status and Contribution of Westin's and Altman's Theories of Privacy." Journal of Social Issues **59**(2): 411-429.
Margulis, S. T. (2003). "Privacy as a Social Issue and Behavioral Concept." Journal of Social Issues **59**(2): 243-261.
MarketLine (2014). "Barnes & Noble, Inc. SWOT Analysis." Barnes & Noble, Inc. SWOT Analysis: 1-10.
Markus, K. A. and D. Borsboom (2013). Frontiers of Test Validity Theory: Measurement, Causation, and Meaning, Routledge
Marsh, H. W. and D. Hocevar (1985). "Application of confirmatory factor analysis to the study of self-concept: First- and higher order factor models and their invariance across groups." Psychological Bulletin **97**(3): 562-582.
Marx, G. T. (2006). Soft Surveillance: The Growth of Mandatory Volunteerism in Collecting Personal information—"hey Buddy Can You Spare a dna?". Surveillance and Security: Technological Politics and Power in Everyday Life. T. Monahan, Routledge**:** 37.
Mason, R. O. (1986). "Four Ethical Issues of the Information Age." MIS Quarterly **10**(1): 5-12.
Massacci, F., M. Prest and N. Zannone (2005). "Using a security requirements engineering methodology in practice: The compliance with the Italian data protection legislation." Computer Standards & Interfaces **27**(5): 445-455.

Maxwell, S. E. and D. A. Cole (2007). "Bias in cross-sectional analyses of longitudinal mediation." Psychological Methods **12**(1): 23-44.
Mayer-Schönberger, V. (2010). "Beyond Privacy, Beyond Rights-Toward a "Systems" Theory of Information Governance." California Law Review **98**(6): 1853-1885.
Mayer-Schonberger, V. and K. Cukier (2013). Big Data: A Revolution That Will Transform How We Live, Work, and Think, John Murray.
McCarthy, J. (2000). "Phenomenal Data Mining." Communications of the ACM **43**(8): 75-79.
McFarlan, F. W. (1988). Editor's Comment. MIS Quarterly, MIS Quarterly & The Society for Information Management. **12:** iii-vi.
McGill, S. and M. Baetz (2011). "Technology use codes of conduct: Is IT a choice between shaping the organizational culture and effective legal enforcement?" Employee Rights & Employment Policy Journal **15**(2): 379-410.
McGuire, T., J. Manyika and M. Chui (2012). "Why Big Data is the New Competitive Advantage." Ivey Business Journal **76**(4): 1-4.
McKechnie, S. (2006). "Integrating intelligent systems into marketing to support market segmentation decisions." Intelligent Systems in Accounting, Finance & Management **14**(3): 117-127.
Meadows, M. and S. Dibb (2012). "Progress in customer relationship management adoption: a cross-sector study." Journal of Strategic Marketing **20**(4): 323-344.
Mesquida, A. L. and A. Mas (2015). "Implementing information security best practices on software lifecycle processes: The ISO/IEC 15504 Security Extension." Computers & Security **48**(0): 19-34.
Messick, S. J. (1989). Validity. Educational measurement. R. L. Linn. New York, American Council on Education and Macmillan**:** 13–103.
Messick, S. J. (1995). "Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning." American Psychologist **50**(9): 741-749.
Milberg, S. J., S. J. Burke, H. J. Smith and A. K. Ernest (1995). "Values, Personal Information Privacy, and Regulatory Approaches. (Cover story)." Communications of the ACM **38**(12): 65-74.
Milberg, S. J., H. J. Smith and S. J. Burke (2000). "Information Privacy: Corporate Management and National Regulation." Organization Science **11**(1): 35-57.
Milne, G. R. (2000). "Privacy and Ethical Issues in Database/Interactive Marketing and Public Policy: A Research Framework and Overview of the Special Issue." Journal of Public Policy & Marketing **19**(1): 1-6.
Milne, G. R. and S. Bahl (2010). "Are There Differences Between Consumers' and Marketers' Privacy Expectations? A Segment- and Technology-Level Analysis." Journal of Public Policy & Marketing **29**(1): 138-149.
Milne, G. R. and M. J. Culnan (2002). "A Longitudinal Analysis of the Privacy Web Sweep Data: Using Marketing Research to Inform Public Policy." The Information Society **18**: 345-359.
Milne, G. R., M. J. Culnan and H. Greene (2006). "A Longitudinal Assessment of Online Privacy Notice Readability." Journal of Public Policy & Marketing **25**(2): 238-249.
Milne, G. R. and A. J. Rohm (2000). "Consumer Privacy and Name Removal Across Direct Marketing Channels: Exploring Opt-In and Opt-Out Alternatives." Journal of Public Policy & Marketing **19**(2): 238-249.
Miller, A. (1972). "Computers, Data Banks and Individual Privacy: An Overview." Columbia Human Rights Law Review **4**: 1-12.
Mingers, J. (2004). "Real-izing information systems: critical realism as an underpinning philosophy for information systems." Information and Organization **14**(2): 87-103.
Mir, R. and A. Watson (2001). "Critical realism and constructivism in strategy research: toward a synthesis." Strategic Management Journal **22**(12): 1169.
Miyazaki, A. and K. Taylor (2008). "Researcher Interaction Biases and Business Ethics Research: Respondent Reactions to Researcher Characteristics." Journal of Business Ethics **81**(4): 779-795.
Miyazaki, A. D. (2008). "Online Privacy and the Disclosure of Cookie Use: Effects on Consumer Trust and Anticipated Patronage." Journal of Public Policy & Marketing **27**(1): 19-33.

Moon, M. J. and S. Bretschneiber (2002). "Does the Perception of Red Tape Constrain IT Innovativeness in Organizations? Unexpected Results from a Simultaneous Equation Model and Implications." Journal of Public Administration Research and Theory **12**(2): 273-292.
Mouncey, P. (2010). Editorial. International Journal of Market Research, World Advertising Research Center Limited. **52:** 285-291.
Mueller, R. O. and G. R. Hancock (2008). Chapter 32: Best practices in Structural Equation Modelling. Best practices in quantitative methods. J. W. Osborne. Thousand Oaks, Calif., Sage Publications.
Mulaik, S. A. (2009). Composite Variables and Linear Transformations. Foundations of Factor Analysis, Second Edition. S. A. Mulaik, Chapman and Hall/CRC**:** 69-92.
Mulaik, S. A., L. R. James, J. Van Alstine, N. Bennett, S. Lind and C. D. Stilwell (1989). "Evaluation of goodness-of-fit indices for structural equation models." Psychological Bulletin **105**(3): 430-445.
Mulligan, D. K. and K. A. Bamberger (2013). "What Regulators Can Do to Advance Privacy Through Design." Communications of the ACM **56**(11): 20-22.
Mulligan, D. K. and A. K. Perzanowski (2007). "The magnificence of the disaster: reconstructing the Sony BMG rootkit incident." Berkeley Technology Law Journal **22**(3): 1157-1232.
Muthén, B., D. Kaplan and M. Hollis (1987). "On structural equation modeling with data that are not missing completely at random." Psychometrika **52**(3): 431-462.
Muthén, B. O. and D. Kaplan (1985). "A comparison of methodologies for the factor analysis of non-normal Likert variables." British Journal of Mathematical and Statistical Psychology **38**(1): 171-189.
Naone, E. (2008). "Who Owns Your Friends?" Technology Review **111**(4): 44-48.
Negash, S. and P. Gray (2008). Business Intelligence. Handbook on Decision Support Systems 2, Springer Berlin Heidelberg**:** 175-193.
Nettleton, E. and I. Turner (2008). "Data protection: Tougher enforcement and increased power for the Information Commissioner?" Journal of Database Marketing & Customer Strategy Management **15**(3): 207-212.
Newman, W. H. and H. W. Wallender Iii (1978). "Managing Not-for-Profit Enterprises." Academy of Management Review **3**(1): 24-31.
Nunan, D. and M. Di Domenico (2013). "Market research and the ethics of big data." International Journal of Market Research **55**(4): 2-13.
O´Hara, K. and N. Shadbolt (2008). The spy in the coffee machine: the end of privacy as we know it. Oxford OX2 7AR, England, Oneworld Publications.
OECD (1980). Annex to the Recommendation of the Council of 23rd September 1980: OECD Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data. O. f. E. C.-o. a. Development.
OECD (1980). OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data.
OECD (2013). The OECD Privacy Framework, ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD).
Ogwueleka, F. N. (2009). "Potential Value of Data Mining for Customer Relationship Marketing in the Banking Industry." Advances in Natural & Applied Sciences **3**(1): 73-78.
Ohlhausen, M. K. (2014). "Privacy Challenges and Opportunities: The Role of the Federal Trade Commission." Journal of Public Policy & Marketing **33**(1): 4-9.
Otto, P. N., A. I. Antón and D. L. Baumer (2006). The ChoicePoint Dilemma: How Data Brokers Should Handle the Privacy of Personal Information. Technical Report, North Carolina State University.
Paas, L. J. (2009). "Database marketing practices and opportunities in a newly emerging African market." Journal of Database Marketing & Customer Strategy Management **16**(2): 92-100.
Paine, C., U.-D. Reips, S. Stieger, A. Joinson and T. Buchanan (2007). "Internet users' perceptions of `privacy concerns' and `privacy actions'." International Journal of Human-Computer Studies **65**(6): 526-536.
Pavlou, P. and D. Gefen (2005). "Psychological contract violation in online marketplaces: antecedents, consequences, and moderating role." Information Systems Research **16**(4): 372–399.
Pavlou, P. A. (2011). "State Of The Information Privacy Literature: Where Are We And Where Should We Go?" MIS Quarterly **35**(4): 977-988.

Pavlou, P. A. (2011). "State of the Information Privacy literature: Where we are now and where should we go?" MIS Quarterly **35**(4): 977-988.
Payne, D. and C. C. Trumbach (2009). "Data mining: proprietary rights, people and proposals." Business Ethics: A European Review **18**(3): 241-252.
Peslak, A. R. (2005). "Internet Privacy Policies: A Review and Survey of the Fortune 50." Information Resources Management Journal **18**(1): 29-41.
Peterson, R. A. (1994). "A Meta-Analysis of Cronbach's Coefficient Alpha." Journal of Consumer Research **21**(2): 381-391.
Pett, M. A., N. R. Lackey and J. J. Sullivan (2003). Assessing the Characteristics of Matrices. Making Sense of Factor Analysis. M. A. Pett, N. R. Lackey and J. J. Sullivan. CA, Thousand Oaks, SAGE Publications Inc.**:** 50-85.
Petty, R. D. (2000). "Marketing Without Consent: Consumer Choice and Costs, Privacy, and Public Policy." Journal of Public Policy & Marketing **19**(1): 42-53.
Pfeffermann, D. and C. R. Rao, Eds. (2009). Sample Surveys: Design, Methods and Applications. Handbook of Statistics 29A. Amsterdam, The Netherlands, North Holland (Elsevier).
Phelps, J., G. Nowak and E. Ferrell (2000). "Privacy Concerns and Consumer Willingness to Provide Personal Information." Journal of Public Policy & Marketing **19**(1): 27-41.
Philip Chen, C. L. and C.-Y. Zhang (2014). "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data." Information Sciences **275**: 314-347.
PI (2009). Privacy International responds to European Commission Consultation on the Privacy Directive Privacy International.
Pierce, M. A. and J. W. Henry (1996). "Computer Ethics: The Role of Personal, Informal and Formal Codes." Journal of Business Ethics **15**: 425-428.
Plosker, G. (2004). Making Money as an Aggregator, Online. **28(2)**.
Popadiuk, S. and C. W. Choo (2006). "Innovation and knowledge creation: How are these concepts related?" International Journal of Information Management **26**(4): 302-312.
Posner, R. A. (1978). "An Economic Theory Of Privacy." REGULATION **MAY/JUNE**.
Purtova, N. (2009). "Property rights in personal data: Learning from the American discourse." Computer Law & Security Review **25**(6): 507-521.
PWC (2014). US cybercrime: Rising risks, reduced readiness: Key findings from the 2014 US State of Cybercrime Survey, PricewaterhouseCoopers LLP; CERT® Division of the Software Engineering Institute at Carnegie Mellon University; CSO magazine; United States Secret Service**:** 1-19.
Raftery, A. E. (1993). Bayesian model selection in structural equation models. Testing structural equation models. K. A. Bollen and J. S. Long. Newbury Park, CA, Sage Publications**:** 163–180.
Rainey, H. G. and Y. H. Chun (2005). Public and Private Management Compared. The Oxford Handbook of Public Management. E. Ferlie, L. E. Lynn and C. Pollitt. Oxford, Oxford University Press**:** 72-102.
Randy. (2015). "Megabytes, Gigabytes, Terabytes... What Are They?", from http://www.whatsabyte.com/.
Rappa, M. (2001). "Business models on the web: Managing the digital enterprise." from http://digitalenterprise.org/models/models.html.
Rappa, M. (2009). "Managing the Digital Enterprise." from http://digitalenterprise.org/about.html.
Rasmussen, K. B. and H. Thimm (2009). "Fact-Based Understanding of Business Survey Non-Response." Electronic Journal of Business Research Methods **7**(1): 83-92.
Reding, V. (2012). "The European data protection framework for the twenty-first century." International Data Privacy Law **2**(3): 119-129.
Refregier, P. and B. Javidi (1995). "Optical image encryption based on input plane and Fourier planerandom encoding." Optics Letters **20**(7): 767-769.
Reidenberg, J. R. (1994). "Setting Standards for Fair Information Practice in the U.S. Private Sector." Iowa Law Review **80**: 497-552.
Reinartz, W., M. Krafft and W. D. Hoyer (2004). "The customer relationship management process: Ist measurement and impact on performance." Journal of Marketing Research **41**(3): 293–305.
Ribak, R. (2007). Privacy Is a Basic American Value: Globalization and the Construction of Web Privacy in Israel. Communication Review, Routledge. **10:** 1-27.

Robinson, M., K. Jones and H. Janicke (2015). "Cyber warfare: Issues and challenges." Computers & Security **49**(0): 70-94.
Robinson, N., H. Graux, M. Botterman and L. Valeri (2009). Review of EU Data Protection Directive: Summary. RAND Europe Technical Report TR-710-ICO. Brussels, Prepared for the Information Commisione's Office
Romanosky, S., D. Hoffman and A. Acquisti (2014). "Empirical Analysis of Data Breach Litigation." Journal of Empirical Legal Studies **11**(1): 74-104.
Roski, J., G. W. Bo-Linn and T. A. Andrews (2014). "Creating Value In Health Care Through Big Data: Opportunities And Policy Implications." Health Affairs **33**(7): 1115-1122.
Rouse, M. (2014). "Data aggregation." WhatIs.com.
Ryker, R., E. Lafleur, B. McManis and K. C. Cox (2002). "Online Privacy Policies: An Assessment of the Fortune E-50." Journal of Computer Information Systems **42**(4): 15.
Ryker, R., E. Latteur, B. McManis and K. C. Cox (2002). "Online Privacy Policies: An Assessment of the Fortune E-50." Journal of Computer Information Systems **Summer**: 15-20.
Samani, R. (2007). "When networks collide." Information Security Technical Report **12**(2): 98-110.
Samiee, S. (1984). "Transnational Data Flow Constraints: A New Challenge for Multinational Corporations." Journal of International Business Studies **15**(1): 141-150.
Samiee, S. (1999). "The internationalization of services: trends, obstacles and issues." Journal of Services Marketing **13**(4/5): 319.
Samuelson, P. (2000). "Privacy As Intellectual Property?" Stanford Law Review **52**(5): 1125.
SAS (2014). Five big data challenges and how to overcome them with visual analytics, SAS Institute Inc.
Scott, D. and A. Bracetti. (2013). "50 Things You Didn't Know About Google." from http://www.complex.com/tech/2013/02/50-things-you-didnt-know-about-google/20-petabytes.
Schein, E. H. (1985). Organizational culture and leadership. San Francisco, Jossey-Bass.
Schermann, M. D., H. D. Hemsen, C. Buchmüller, T. Bitter, H. P. D. Krcmar, V. P. D. Markl and T. P. D. Hoeren (2014). "Big Data." Business & Information Systems Engineering **6**(5): 261-266.
Schillewaert, N., F. Langerak and T. Duhamel (1998). "Non-probability sampling for WWW surveys: a comparison of methods." Journal of the Market Research Society **40**(4): 307-322.
Schnieier, B. (2002). Computer Security: It's the Economics, Stupid. Workshop on Economics and Information Security. University of California, Berkeley, University of Cambridge (UK).
Schwaig, K. S., G. C. Kane and V. C. Storey (2006). "Compliance to the fair information practices: How are the Fortune 500 handling online privacy disclosures?" Information & Management **43**(7): 805-820.
Schwartz, P. M. (1994). "European Data Protection Law and Restrictions on International Data Flows." Iowa Law Review **80**(1994-95): 471-496.
Sebor, J. (2007). "TOO MUCH pork for JUST ONE fork." CRM Magazine **11**(5): 28-31.
Seddon, P. B. and R. Scheepers (2012). "Towards the improved treatment of generalization of knowledge claims in IS research: drawing general conclusions from samples." European Journal of Information Systems **21**(1): 6-21.
Shaffer, G. (1999). "The Power of EU Collective Action: The Impact of EU Data Privacy Regulation on US Business Practice." European Law Journal **5**(4): 419-437.
Shaffer, G. (2000). "Globalization and Social Protection: The Impact of EU and International Rules in the Ratcheting up of US Data Privacy Standards." Yale Journal of International Law **25**: 1-88.
Shalhoub, Z. K. (2009). "Analysis of Industry-Specific Concentration of CPOs in Fortune 500 Companies." Communications of the ACM **52**(4): 136-141.
Sheehan, K. B. (2005). "In Poor Health: An Assessment of Privacy Policies at Direct-to-Consumer Web Sites." Journal of Public Policy & Marketing **24**(2): 273-283.
Sheskin, D. J. (2003). Introduction. Handbook of Parametric and Nonparametric Statistical Procedures, Chapman and Hall/CRC.
Shi-Ming, H., L. Chia-Ling and K. Ai-Chin (2006). "Balancing performance measures for information security management: A balanced scorecard framework." Industrial Management & Data Systems **106**(1/2): 242-255.

Shires, K. (2011). "How to protect data in times of change? A review of the EU Commission's and ICO's response to updating the law on data protection." Journal of Database Marketing & Customer Strategy Management **18**(1): 65-68.
Shook, C. L., D. J. Ketchen, Jr., G. T. M. Hult and K. M. Kacmar (2004). "An Assessment of the Use of Structural Equation Modeling in Strategic Management Research." Strategic Management Journal **25**(4): 397-397+.
Siegel, S. (1956). Nonparametric statistics for the behavioral sciences. London, McGraw-Hill.
Simitis, S. (1994). "From the Market to the Polis: The EU Directive on the Protection of Personal Data." Iowa Law Review **80**: 445-470.
Siponen, M., S. Pahnila and M. A. Mahmood (2010). "Compliance with Information Security Policies: An Empirical Investigation." Computer **43**(2): 64-71.
Six, P. (1998). The future of privacy - Volume 1: Private life and public policy. London, Demos.
Skrondal, A. and S. Rabe-Hesketh (2004). Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models. C&H/CRC Monographs on Statistics & Applied Probability, Chapman and Hall/CRC
Smith, D. and K. Langfield-Smith (2004). "Structural Equation Modeling in Management Accounting Research: Critical Analysis and Opportunities." Journal of Accounting Literature **23**: 49.
Smith, H. J. (1993). "Privacy Policies and Practices: Inside the Organizational Maze." Communications of the ACM **36**(12): 105-122.
Smith, H. J., T. Dinev and H. Xu (2011). "Information Privacy Research: An Interdisciplinary Review." MIS Quarterly **35**(4): 980-A927.
Smith, H. J., S. J. Milberg and S. J. Burke (1996). "Information Privacy: Measuring Individuals' Concerns About Organizational Practices." MIS Quarterly **20**(2): 167-196.
Smith, M. L. (2006). "Overcoming theory-practice inconsistencies: Critical realism and information systems research." Information and Organization **16**(3): 191-211.
Snijkers, G., G. Haraldsen, J. Jones and D. Willimack (2013). Designing and Conducting Business Surveys. Hoboken (New Jersey), JohnWiley& Sons, Inc.
Solove, D. (2006). "A taxonomy of privacy." University of Pennsylvania Law Review **154**(3): 477.
Solove, D. J. (2004). The digital person: technology and privacy in the information age. New York, New York University Press.
Son, J.-Y. and S. S. Kim (2008). "Internet Users' Information Privacy-Protective Responses: A Taxonomy and a Nomological Model." MIS Quarterly **32**(3): 503-529.
Sophos. (2010, 19 July 2010). "Data protection laws too lax, Sophos survey reveals." Retrieved 20 June 2011, from http://www.sophos.com/en-us/press-office/press-releases/2010/07/data-protection.aspx.
Sovern, J. (1999). "Opting In, Opting Out, or No Options at all: the fight for control of personal information." Washington Law Review **74**: 1033-1118.
Spangler, W. E., M. Gal-Or and J. H. May (2003). "USING DATA MINING TO PROFILE TV VIEWERS." Communications of the ACM **46**(12): 66-72.
Spanos, G. and L. Angelis (2016). "The impact of information security events to the stock market: A systematic literature review." Computers & Security **58**: 216-229.
Spearman, C., C. (1910). "Correlation calculated from faulty data." British Journal of Psychology **3**: 271–295.
Spears, J. L. and H. Barki (2010). "User Participation in Information Systems Security Risk Management." MIS Quarterly **34**(3): 503-A505.
Stalder, F. (2002). "Privacy is not the antidote to surveillance." Surveillance & Society **1**(1): 120-124.
Stalder, F. (2011). "Autonomy beyond Privacy? A Rejoinder to Colin Bennett." Surveillance & Society **8**(4): 508-512.
Stanley, J. (2004). The Surveillance-Industrial Complex. How the American Government Is Conscripting Businesses and Individuals in the Construction of a Surveillance Society. A. C. L. U. (ACLU). New York.
Stata. (2015). "mvtest normality — Multivariate normality tests." Retrieved 28/05/2014, from http://www.stata.com/manuals13/mvmvtestnormality.pdf.

Staten, M. E. and F. H. Cate (2003). "The impact of opt-in privacy rules on retail credits markets: a case study of MBNA." Duke Law Journal **52**(4): 745-786.
Stavridis, J. G. and E. C. Parker Iii (2012). "Sailing the Cyber Sea." Joint Force Quarterly(65): 61-67.
Steiger, J. H. and J. C. Lind (1980). Statistically-based tests for the number of common factors. Annual Spring Meeting of the Psychometric Society. Iowa City.
Stevens, B. (2008). "Corporate Ethical Codes: Effective Instruments for Influencing Behavior." Journal of Business Ethics(78): 601-605.
Stone, E. F., D. G. Gardner, H. G. Gueutal and S. McClure (1983). "A Field Experiment Comparing Information-Privacy Values, Beliefs, and Attitudes Across Several Types of Organizations." Journal of Applied Psychology **68**(3): 459-468.
Storey, V. C., G. C. Kane and K. S. Schwaig (2009). "The Quality of Online Privacy Policies: A Resource-Dependency Perspective." Journal of Database Management **20**(2): 19-37.
Strahilevitz, L. J. (2008). "Privacy versus Antidiscrimination." University of Chicago Law Review **75**(1): 363-381.
Straub, D. W. and R. J. Welke (1998). "Coping with Systems Risk: Security Planning Models for Management Decision Making." MIS Quarterly **22**(4): 441-469.
Tajahuerce, E. and B. Javidi (2000). "Encrypting three-dimensional information with digital holography." Applied Optics **39**(35): 6595-6601.
Tanaka, H., K. Matsuura and O. Sudoh (2005). "Vulnerability and information security investment: An empirical analysis of e-local government in Japan." Journal of Accounting and Public Policy **24**(1): 37-59.
Tang, Z., Y. U. Hu and M. D. Smith (2008). "Gaining Trust Through Online Privacy Protection: Self-Regulation, Mandatory Standards, or Caveat Emptor." Journal of Management Information Systems **24**(4): 153-173.
Tankard, C. (2012). "Big data security." Network Security **2012**(7): 5-8.
Taylor, D. (1996). "Aggregating data makes it useful: real-world lessons.(Inside the Firewall) (Technology Information)(Column)." InfoWorld **18**(48): 12(11).
Taylor, H. (1999). "Does internet research work?" International Journal of Market Research **42**(1): 51-63.
Tene, O. (2010). "Privacy: The New Generations." International Data Privacy Law: 1-13.
Tene, O. (2012). "Privacy in the Age of Big Data: A Time for Big Decisions." Stanford Law Review **64**: 63-69.
Tene, O. and J. Polenetsky (2012). "To Track or Do Not Track: Advancing Transparency and Individual Control in Online Behavioral Advertising." Minnesota Journal of Law, Science & Technology **13**: 281-358.
TheEconomist (2010). Data, data everywhere. Managing information.
Thiel, S. (2008). Reed Elsevier to Buy ChoicePoint for $3.5 Billion (Update10) Bloomberg.
Thomas, H. D. and D. B. Jeffrey (2004). "Enterprise systems and the supply chain." Journal of Enterprise Information Management **17**(1): 8-19.
Thomas, R. E. and V. G. Maurer (1997). "Database Marketing Practice: Protecting Consumer Privacy." Journal of Public Policy & Marketing **16**(1): 147-155.
Thompson, B. (2004). Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications, American Psychological Association (APA).
Tirial. (2009). "The TJX data breach case." Retrieved 18 April 2011, from http://www.bukisa.com/articles/107044_the-tjx-data-breach-case.
Trepte, S. and L. Reinecke, Eds. (2011). Privacy Online: Perspectives on Privacy and Self-Disclosure in the Social Web. Berlin, Springer.
Tsai, J., S. Egelman, L. Cranor and A. Acquisti (2008). The effect of online privacy information on purchasing behavior: An experimental study. The 6th Workshop on the Economics of Information Security (WEIS)**:** 33.
Turow, J. and N. Draper (2012). 2.1.b Advertising's new surveillance ecosystem. Handbook of Surveillance Studies. K. Ball, K. D. Haggerty and D. Lyon. New York, Routledge**:** 133-140.
Tuten, T. L., D. J. Urban and M. Bosnjak (2002). Internet Surveys and Data Quality: A review. Online Social Sciences. B. Batinic, U.-D. Reips, M. Bosnjak and A. Werner, Hogrefe Publishing**:** 7-28.

UN (2004). Chapter 6: Protecting Privacy Rights in an Online World. United Nation Conferece on Trade and Development. New York, Geneva.
Urbach, N. and F. Ahlemann (2010). "Structural Equation Modeling in Information Systems Research Using Partial Least Squares." JITTA : Journal of Information Technology Theory and Application **11**(2): 5-39.
US-EU (2000). US-EU Safe Harbor Framework Documents. Export.gov.
US-Senate (2013). A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes. STAFF REPORT FOR CHAIRMAN ROCKEFELLER. O. O. O. A. I. M. STAFF, United States Senate. **December 18, 2013:** 36.
USC (1974). The Privacy Act of 1974. National Archives. **Pub. L. No. 93-579, 5 U.S.C. §552a**.
Vaidya, J. and V. Atluri (2007). Chapter 6. Privacy, Profiling, Targeted Marketing, and Data Mining. Digital Privacy: Theory, Technologies, and Practices. A. Publications, Sabrina De Capitani di Vimercati, Stefanos Gritzalis, Costas Lambrinoudakis, Alessandro Acquisti**:** 117-131.
Vail, M. W., J. B. Earp and A. L. Antón (2008). "An Empirical Study of Consumer Perceptions and Comprehension of Web Site Privacy Policies." IEEE Transactions on Engineering Management **55**(3): 442-454.
van Blarkom, G. W., J. J. Borking and J. G. E. Olk, Eds. (2003). Handbook of Privacy and Privacy-Enhancing Technologies: The case of Intelligent Software Agents. PISA Consortium. The Netherlands, The Hague.
van der Sloot, B. (2014). "Do data protection rules protect the individual and should they? An ssessment of the proposed General Data Protection Regulation." International Data Privacy Law **4**(4): 307-325.
van Dijck, J. (2014). "Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology." Surveillance & Society **12**(2).
Varian, H. R. (1985). "Price Discrimination and Social Welfare." American Economic Review **75**(4): 870-875.
Varian, H. R. (2010). "Richard T. Ely Lecture: Computer Mediated Transactions." American Economic Review **100**(2): 1-10.
Velu, R. and G. M. Naidu (2009). Survey Sampling Methods in Marketing Research: AReview of Telephone, Mall Intercept, Panel, andWeb Surveys. Sample Surveys: Design, Methods and Applications, Vol. 29A. D. Pfeffermann and C. R. Rao, Elsevier**:** 511-538.
Venaik, S., D. F. Midgley and T. M. Devinney (2004). "A new perspective on the integration-responsiveness pressures confronting multinational firms." Management International Review **44**(1): 15–48.
Venaik, S., D. F. Midgley and T. M. Devinney (2005). "Dual paths to performance: the impact of global pressures on MNC subsidiary conduct and performance." Journal of International Business Studies **36**(6): 655–675.
Verhagen, T., B. van Den Hooff and S. Meents (2015). "Toward a Better Use of the Semantic Differential in IS Research: An Integrative Framework of Suggested Action." Journal of the Association for Information Systems **16**(2): 108-143.
Voelpel, S. C., M. Dous and T. H. Davenport (2005). "Five steps to creating a global knowledge-sharing system: Siemens' ShareNet." Academy of Management Executive **19**(2): 9-23.
Volpentesta, A. P., S. Ammirato and R. Palmieri (2011). "Investigating effects of security incident awareness on information risk perception." International Journal of Technology Management **54**(2): 304-320.
Walker, R. M. (2014). "Internal and External Antecedents of Process Innovation: A review and extension." Public Management Review **16**(1): 21-44.
Wang, C.-Y. (2012). "A knowledge network production: Ten years of information security research." African Journal of Business Management **6**(1): 213-221.
Wang, X., Y. L. Yin and H. Yu (2005). Finding Collisions in the Full SHA-1. Advances in Cryptology – CRYPTO 2005: 25th Annual International Cryptology Conference, Santa Barbara, California, USA, August 14-18, 2005. Proceedings. V. Shoup. Berlin, Heidelberg, Springer Berlin Heidelberg**:** 17-36.
Warren, A., R. Bayley, C. Bennett, A. Charlesworth, R. Clarke and C. Oppenheim (2008). "Privacy Impact Assessments: International experience as a basis for UK Guidance." Computer Law & Security Review **24**(3): 233-242.
Wasserman, L. (2004). All of Statistics: A Concise Course in Statistical Inference, Springer**:** 442.

Wasserman, L. (2006). All of Nonparametric Statistics. New York, US, Springer**:** 280.
Watson, H. J., B. H. Wixom, J. A. Hoffer, R. Anderson-Lehman and A. M. Reynolds (2006). "Real-Time Business Intelligence: Best Practices at Continental Airlines." Information Systems Management **23**(1): 7-18.
Wei-Hsiu, W. and L. Woo-Tsong (2014). "Development Trends and Strategy Planning in Big Data Industry." Contemporary Management Research **10**(3): 203-213.
Weitzner, D., H. Abelson, T. Berners-Lee, J. Feigenbaum, J. Hendler and G. Sussman (2008). "Information accountability." Communications of the ACM **51**(6): 82-87.
Weitzner, D. J. A., Harold; Berners-Lee, Tim; Hanson, Chris; Hendler, James; Kagal, Lalana; McGuinness, Deborah L.; Sussman, Gerald Jay; Waterman, K. Krasnow (2006). Transparent Accountable Data Mining: New Strategies for Privacy Protection. Massachusetts Institute of Technology Computer Science and Artificial Intelligence Laboratory. MIT. **MIT-CSAIL-TR-2006-007**.
Westin, A. F. (1967). Privacy and freedom. New York, Athenäum.
Westin, A. F. (2003). "Social and Political Dimensions of Privacy." Journal of Social Issues **59**(2): 431-453.
Weston, D. J., D. J. Hand, N. M. Adams, C. Whitrow and P. Juszczak (2008). "Plastic card fraud detection using peer group analysis." Advances in Data Analysis and Classification **2**: 45-62.
Whitley, E. A. (2009). "Informational privacy, consent and the "control" of personal data." Information Security Technical Report **14**(3): 154-159.
Whitman, J. Q. (2004). "The Two Western Cultures of Privacy: Dignity versus Liberty." The Yale Law Journal **113**(6): 1151-1221.
Wilcox, J. B., R. D. Howell and E. Breivik (2008). "Questions about formative measurement." Journal of Business Research **61**(12): 1219-1228.
Winklhofer, H. M. and A. Diamantopoulos (2002). "Managerial evaluation of sales forecasting effectiveness: A MIMIC modeling approach." International Journal of Research in Marketing **19**(2): 151-166.
Witt, P. and V. Rode (2005). "Corporate brand building in start-ups." Journal of Enterprising Culture **13**(3): 273–294.
WMI (2014). Sony Corporation (6758) : Company Profile and SWOT Analysis. World Market Intelligence. London, Progressive Media Group**:** 1-45.
Wood, D. M. and K. Ball (2013). "Brandscapes of control? Surveillance, marketing and the co-construction of subjectivity and space in neo-liberal capitalism." Marketing Theory **13**(1): 47-67.
Wood, R. A. (2000). "Market Microstructure Research Databases: History and Projections." Journal of Business & Economic Statistics **18**(2): 140-145.
Wood, Z. and T. Lyons (2010). Clubcard couple head for checkout at Tesco. Pair who invented ground-breaking loyalty card scheme bid quiet farewell to key client and majority owner, Tesco. guardian.co.uk.
Wry, T. (2009). "Does Business and Society Scholarship Matter to Society? Pursuing a Normative Agenda with Critical Realism and Neoinstitutional Theory." Journal of Business Ethics **89**(2): 151-171.
Wu, L. and E. Brynjolfsson (2012). The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales. Rochester, Social Science Research Network.
Wynne, B. (1975). "The rhetoric of consensus politics: a critical review of technology assessment." Research Policy **4**(2): 108-158.
Xavier, M. J., A. Srinivasan and A. Thamizhvanan (2011). "Use of analytics in Indian enterprises: an exploratory study." Journal of Indian Business Research **3**(3): 168-179.
Yau-De, W., Y. Chyan and W. Kuei-Ying (2012). "Comparing Public and Private Employees' Job Satisfaction and Turnover." Public Personnel Management **41**(3): 557-573.
Zarsky, T. Z. (2002). "Mine Your Own Business: Making the Case for the Implications of the Data Mining of Personal Information in the Forum of Public Opinion." Yale JL & Tech. **5**: 1.
Zicari, R. V. (2013). Chapter 3. Big Data: Challenges and Opportunities. Big Data Computing. R. Akerkar, Chapman and Hall/CRC**:** 103-128.
Zureik, E., L. L. Harling Stalker, E. Smith, D. Lyon and Y. E. Chan, Eds. (2010). Surveillance, Privacy, and the Globalization of Personal Information: International Comparisons. Montreal, Canada, McGill-Queen's University Press.